DOCUMENT RESUME

ED 385 574 TM 024 012

AUTHOR Carlton, Sydell T.; Harris, Abigail M.
TITLE Characteristics Associated with Differential Item Functioning on the Scholastic Aptitude Test: Gender and Majority/Minority Group Comparisons.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RR-92-64
PUB DATE Nov 92
NOTE 192p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC08 Plus Postage.
DESCRIPTORS American Indians; Asian Americans; Black Students; Comparative Analysis; Ethnic Groups; High Schools; High School Students; Hispanic Americans; *Item Bias; *Minority Groups; *Racial Differences; *Sex Differences; *Test Items; White Students
IDENTIFIERS *Mantel Haenszel Procedure; *Scholastic Aptitude Test; Test Specifications

ABSTRACT

The purpose of the study was to investigate whether selected test and item characteristics in the Scholastic Aptitude Test (SAT) are associated with unexpected differential item functioning (DIF) for males and females and for majority and minority group members (i.e., White performance compared with Black, Asian American, Hispanic American, and American Indian performance). Six forms of the SAT, with 379,896 examinees in all, each containing verbal and mathematics sections and the Test of Standard Written English (TSWE), were used. Findings from previous studies, test specifications, and suggestions from experts led to the identification of more than 100 a priori item coding categories. Each SAT item was coded accordingly, by type, content, and format. The Mantel-Haenszel procedure was used to provide an index of DIF for each reference (males and Whites) and focal group (females and ethnic groups other than White) comparison. The study identifies and reports on patterns of differential performance and section-specific differences for gender and ethnic groups. Three appendixes contain the coding categories for verbal and mathematics sections and the TSWE. Study findings are reported in 38 tables. (Contains 29 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

RR-92-64

ED 385 574

RESEARCH REPORT

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
H. I. BRAUN
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)"

CHARACTERISTICS ASSOCIATED WITH DIFFERENTIAL ITEM FUNCTIONING ON THE SCHOLASTIC APTITUDE TEST: GENDER AND MAJORITY/MINORITY GROUP COMPARISONS

Sydell T. Carlton
Abigail M. Harris

Educational Testing Service
Princeton, New Jersey
November 1992

TM024012

BEST COPY AVAILABLE

CHARACTERISTICS ASSOCIATED WITH DIFFERENTIAL ITEM FUNCTIONING ON THE SCHOLASTIC APTITUDE TEST: GENDER AND MAJORITY/MINORITY GROUP COMPARISONS

Sydell T. Carlton

Educational Testing Service

and

Abigail M. Harris

Fordham University

Copyright © 1992 by Educational Testing Service, Princeton, N.J. All rights reserved.

TABLE OF CONTENTS

	Page
INTRODUCTION	1
PROCEDURE	5
Test Forms	5
Description of the SAT	6
Coding Categories	7
Analyses	9
Item Analyses	9
Mantel-Haenszel	9
Analysis of Variance	11
RESULTS	12
Overall Performance	13
Item Category Results	18
Points Tested	19
Verbal Scale	19
Mathematics Scale	21
TSWE	25
Item Content	25
Subject Matter	26
Technical/Non-Technical	30
Concrete/Abstract	31
Latinate Language	32
Homographs	32
Parts of Speech	33
Emotive or Controversial Material	34
Minority and Gender Reference	35
Item Format	38
Length	38
Vertical/Wraparound Response Format	39
First Appearance	40
Clues to Answer	41
Other Format Variables--Verbal	43
Other Format Variables--Mathematics	43
Other Format Variables--TSWE	45
Miscellaneous Format Variables	46
DISCUSSION and SUMMARY	47
Highlights--Gender Differences	48
Verbal and TSWE	48
Mathematics	50

Highlights--White/Racial/Ethnic Background Comparisons 52
Verbal and TSWE 52
Mathematics 57

Implications 58
Research Directions 61

REFERENCES 64

APPENDICES 67
A Coding Categories: Verbal 67
B Coding Categories: Mathematics 86
C Coding Categories: Test of Standard Written English 103

TABLES 108

TABLES

Table 1 *Summary of SAT and TSWE Test Form Administration Dates and Sample Sizes*

Table 2 *Summary of Mean Formula Score Differences Between Reference and Focal Groups in Standard Deviation Units*

Table 3 *Average DIF* Values for Reference/Focal Comparison Groups*

Table 4 *Number and percent of SAT Items by Section and Reference/Focal Group Comparison that Evidence DIF Greater Than 1.0 or Less Than 1.0*

Table 5 *Number and Percent of SAT Items by Section and Reference/Focal Group Comparison that Evidence DIF Greater Than 1.5 or Less Than 1.5*

Table 6 *Means and Standard Deviation for Item Categories on the Analogies Section of the SAT-V: Points Tested*

Table 7 *Means and Standard Deviation for Item Categories on the Antonyms Section of the SAT-V: Points Tested*

Table 8 *Means and Standard Deviation for Item Categories on the Reading Comprehension Section of the SAT-V: Points Tested*

Table 9 *Means and Standard Deviation for Item Categories on the SAT-M: Points Tested*

Table 10 *Means and Standard Deviations for Item Categories on the SAT-M: Points Tested (Subcontent Areas)*

Table 11 *Means and Standard Deviations for Item Categories on the SAT-M: Points Tested (Spatial/Visual Factor)*

Table 12 *Means and Standard Deviations for Item Categories on the TSWE: Points Tested*

Table 13 *Means and Standard Deviations for Item Content Categories on the SAT-V: Sentence Completions, Analogies, and Antonyms*

Table 14 *Means and Standard Deviations for Item Content Categories on the SAT-V: Reading Comprehension*

Table 15 *Means and Standard Deviations for Item Content Categories on the TSWE*

Table 16 *Means and Standard Deviations for Item Content Categories on the SAT-M*

Table 17 *Means and Standard Deviations for Item Content Categories on the SAT-V: Technical/Non-Technical*

Table 18 *Means and Standard Deviations for Item Content Categories on the Analogies Section of the SAT-V: Concrete/Abstract*

Table 19 *Means and Standard Deviations for Item Content Categories on the Analogies Section of the SAT-V: Latinate Language*

Table 20 *Means and Standard Deviations for Item Content Categories on the Analogies and Antonyms Sections of the SAT-V: Homographs*

Table 21 *Means and Standard Deviations for Item Content Categories on the SAT-V: Parts of Speech*

Table 22 *Means and Standard Deviations for the Emotive Item Content Category on the SAT-V and TSWE*

Table 23 *Means and Standard Deviations for Item Content Categories on the SAT-V and TSWE: Controversial and Socially Relevant*

Table 24 *Means and Standard Deviations for the Minority Stimulus Item Content Categories on the Sentence Completions and Reading Comprehension Sections of the SAT-V*

Table 25 *Means and Standard Deviations for the Gender Reference Item Content Categories on the Sentence Completions and Reading Comprehension Sections of the SAT-V*

Table 26 *Means and Standard Deviations for the Gender Reference and Minority Stimulus Item Content Categories on the TSWE*

Table 27 *Means and Standard Deviations for the People Reference Item Content Categories on the Analogies and Antonyms Sections of the SAT-V*

Table 28 *Means and Standard Deviations for the Gender Reference and Minority Stimulus Item Content Categories on the SAT-M*

Table 29 *Summary of Analysis of Variance Results for the Item Format Categories Dealing with Length on the SAT-V and TSWE*

Table 30 *Summary of Analysis of Variance Results for the Item Format Categories Dealing with Length on the SAT-M*

Table 31 *Means and Standard Deviations for the Option Format Item Category on the SAT*

Table 32 *Means and Standard Deviations for Item Format Categories Dealing with First Appearance and First Reappearance of an Item Type on the SAT-V and TSWE*

Table 33 *Means and Standard Deviations for Item Format Categories Dealing with First Appearance and First Reappearance of an Item Type on the SAT-M*

Table 34 *Means and Standard Deviations for Item Format Categories on the Analogies Section of the SAT-V: Clues to the Answer*

Table 35 *Means and Standard Deviations for Item Format Categories on the Sentence Completions and Reading Comprehension Sections of the SAT-V: Clues to the Answer*

Table 36 *Means and Standard Deviations for Item Format Categories of the SAT-M*

Table 37 *Means and Standard Deviations for Item Format Categories on the TSWE*

Table 38 *Summary of Analysis of Variance Results for Selected Item Format Categories on the SAT*

ABSTRACT

The purpose of the study was to investigate whether selected test and item characteristics in the SAT are associated with unexpected differential item functioning (DIF) for males and females and for majority and minority group members (i.e., White performance compared with Black, Asian-American, Hispanic, and American Indian performance). Six forms of the SAT that were administered relatively recently were used. Each of the six forms consists of Verbal sections (containing Analogies, Antonyms, Sentence Completions, and Reading Comprehension), Mathematics sections (containing Regular Mathematics problem-solving and Quantitative Comparisons), and the Test of Standard Written English (TSWE) (containing Usage and Sentence Correction). Findings from previous studies, test specifications, and suggestions offered by test development experts led to the identification of more than one hundred *a priori* item coding categories, and each SAT item was coded accordingly. Items were coded with regard to type, content, and format.

The Mantel-Haenszel procedure was used to provide an index of differential item functioning (DIF) for each reference/focal group comparison. Females, Asian-Americans, Blacks, Hispanics, and American Indians were the focal groups; males and Whites were the reference groups. With the exception of the American Indians, each focal group was represented in sufficient numbers on each test form to lead to meaningful interpretation of data. For each item category, one-way analyses of variance were computed using as the dependent variable the Mantel-Haenszel DIF values. Analyses were run separately for each reference/focal group investigated.

The study reports on patterns of differential performance across SAT sections as well as on section-specific differences. Patterns across several ethnic groups and between two gender groups, as well as results specific to each of the groups, are identified. The report addresses the following questions:

- Are there unexpected group differences associated with the points tested (e.g., percentages, verb forms)?
- What aspects of item content (e.g., subject matter, gender and ethnic references, level of language) are associated with unexpected group differences?
- Are there elements of test or item format (e.g., length of stem, formatting or location of the stem or options) that are associated with unexpected group differences?

CHARACTERISTICS ASSOCIATED WITH DIFFERENTIAL ITEM FUNCTIONING ON THE SCHOLASTIC APTITUDE TEST: GENDER AND MAJORITY/MINORITY GROUP COMPARISONS

Sydell T. Carlton and Abigail M. Harris

Educational Testing Service Fordham University

Introduction

The Scholastic Aptitude Test (SAT) is recognized as an important tool in the college selection and admissions process. Educational Testing Service (ETS), which develops and administers the SAT for the College Board, is sensitive both to the critical role that the test plays and to the diversity of individuals who take the test, and its test developers and researchers are committed to reviewing the test and test performance to ensure that the test is fair for examinees regardless of ethnic or gender group membership. In recent years several strategies have been used routinely to detect and eliminate possible favoritism in items on the SAT (Carlton & Marco, 1982). Every ETS test must undergo a sensitivity review process prior to administration, including its pretesting administration (Hunter & Slaughter, 1980). All SAT items are pretested (or tried out prior to their use), and items that the data suggest are unexpectedly more difficult for one group than for another are reevaluated for possible elimination from the item pool. Following test administrations, retrospective investigations are done to evaluate item performance; analyses of group differences (known as analyses for Differential Item Functioning, or DIF) are part of this process. Further,

results of these ongoing efforts have given rise to additional exploratory studies of possible group differences in such areas as test speededness (Wild & Durso, 1979) analogy problem-solving strategies (Freedle, Kostin, & Schwartz, 1987), error choices (Donlon, 1982, cited in Clark and Grandy, 1984), and the influence of language proficiency on performance (Bleistein & Wright, 1987).

Despite continuing efforts, however, there is some evidence to suggest that the SAT underpredicts for females (Clark & Grandy, 1984) and overpredicts for selected other groups when a majority or common regression line is used to predict college freshman grade point average (GPA) (Linn, 1973). This has led critics to suggest that group mean differences in performance may be reflecting test bias or group differences in background rather than actual differences in the construct being measured (Rosser, 1987). One way to address this concern and possibly to gain insights into improving GPA prediction is to systematically explore the ways in which several groups that have been matched on overall performance differ in terms of performance on individual items. When patterns of differential item performance are identified, they can be reviewed to determine whether they reflect valid differences between groups or whether they may be an indication of bias. For example, differential item functioning on a subset of items measuring facility with mathematics problem-solving may reflect real group differences in mathematics problem-solving skills or it may indicate that the items are measuring something other than or in addition to mathematics problem-solving. When one group of examinees is in fact better at solving mathematics problems than another group, it could be due to group differences in ability to solve

mathematics problems or it could reflect group differences in such factors as prior participation in higher-level mathematics courses or mathematical game activities; in both of these cases, the differences in item functioning reflect true group differences and as such are not considered to be indicative of test bias. Alternatively, if an item is intended to measure mathematics problem-solving but the content of the problem in which the mathematics is embedded is such that there are group differences in familiarity with the terminology (e.g., references to areas to which one group is likely to have had much more or much less exposure than another group), differential item functioning may in fact be indicative of bias. Some examinees may be more successful on the item not because they are better at solving mathematics problems or because they took more mathematics courses but rather because they have the unfair advantage of familiarity with the surrounding content.

Similarly, there may be item formats that favor one group or another. If all of the mathematics problem-solving items are posed using a particular item format (e.g., word problems), one group may appear to be better at problem-solving than another group when, in fact, the differences may be due to different ways or styles of interpreting or reacting to the test stimuli.

Becoming aware of patterns of differential item performance that suggest actual group differences has implications for policy-makers and practitioners, including educators (e.g., teachers, curriculum developers) and state and local boards of education as they make decisions about preparation of students for college. Identifying items that exhibit differential item functioning based on group membership is useful to test developers and test policymakers as they evaluate or reconsider the value or relevance of different kinds of

test items to the construct being measured and to predicting success in college.

This study is the most comprehensive of its kind to date. Its purpose was to investigate whether selected test or item characteristics in the SAT are associated with unexpected differential item functioning (DIF) for males and females and for majority and minority group members (i.e., White performance compared with Black, Asian-American, Hispanic, and American Indian performance). Included are SAT Verbal sections (containing Analogies, Antonyms, Sentence Completions, and Reading Comprehension), SAT Mathematics sections (containing Regular Mathematics problem-solving and Quantitative Comparisons), and the Test of Standard Written English (TSWE) section (containing Usage and Sentence Correction) from six forms of the SAT. Unlike past studies, which have been more limited in scope, this study focuses on patterns of differential performance across SAT sections as well as on section-specific differences. Patterns across several ethnic groups and between the two gender groups, as well as findings specific to each of the groups, are identified.

In this report, the following questions will be addressed:

* Are there unexpected group differences associated with the points being tested (e.g., percentages, verb forms)?
* What aspects of item content (e.g., subject matter, gender and ethnic references, level of language) are associated with unexpected group differences?
* Are there elements of test or item format (e.g., length of stem, formatting or location of the stem or options) that are associated with unexpected group differences?

Procedure

Six forms of the SAT that were relatively recently administered were selected for study. Past studies have tended to use fewer forms. The inclusion of six forms made it possible to increase the sample sizes of groups and to increase the item pool. Findings from previous studies, test specifications, and suggestions offered by test development experts led to the identification of more than a hundred a priori item coding categories, and each SAT item was coded accordingly. Items were coded with regard to type, content, and format.

The Mantel-Haenszel procedure was used to provide an index of differential item functioning (DIF) for each reference/focal group comparison. Females, Asian-Americans, Blacks, Hispanics, and American Indians were the focal groups; males and Whites were the reference groups. For each item category, one-way analyses of variance were computed using as the dependent variable the Mantel-Haenszel DIF values. Analyses were run separately for each reference/focal group investigated.

Test Forms

The six SAT forms (see Table 1) were selected from those administered

---

Insert Table 1 about here

---

over a three-year period, from January 1983 to December 1985. They satisfied the conditions of being recent, of covering a variety of test dates, and of

having DIF data available. With the exception of American Indians, each focal group was represented in sufficient numbers on each form to lead to meaningful interpretation of data. In the aggregate, the six SAT forms yielded data for 181,228 males and 198,668 females; and for 279,814 Whites, 16,073 Asian-Americans, 40,184 Blacks, 13,624 Hispanics, and 3041 American Indians. The six TSWE forms that were selected were all administered in 1985. For each, sample sizes were adequate and DIF data were available for the identified reference/focal group comparisons. Samples in all cases were limited to high school juniors and seniors for whom English was self-reported to be the best language. This study does not investigate the performance of examinees who reported that English is not their best language. In past investigations with Hispanics (e.g., Alderman, 1982; Schmitt, 1988) and Asian-Americans (Bleistein and Wright, 1987; Dorans and Kulick, 1983, 1986) language proficiency was an important factor in extraneous variability in test scores and differential item functioning. Further, ambiguity of the identifying question for American Indians (which led some or perhaps many Whites to respond as American Indians), coupled with the very low number of American Indians, led the investigators to conclude that the data were not likely to be reliable; consequently, the White/American Indian comparison was dropped in this study.

Description of the SAT

Each form of the SAT consists of five operational (scored) thirty-minute sections: two Verbal (V), two Mathematics (M), and one Test of Standard Written English (TSWE). The two Verbal sections combined consist of 85 items, which cover four item types: 25 Reading Comprehension, 25 Antonyms, 20 Analogies, and 15 Sentence Completions. Of these Verbal items, the 25 Reading

Comprehension items are in six sets; that is, they are drawn from six reading passages. The other 60 items are "discrete"; that is, they are separate, or unlinked. The two Mathematics sections combined consist of 60 items, which cover two item types: 40 "Regular Mathematics," or problem-solving, items and 20 Quantitative Comparison items. The 50-item TSWE section consists of two item types: 35 Usage items (which require the recognition of where in a sentence, if at all, errors occur) and 15 Sentence Correction items (which require the recognition of the best way of writing a given sentence). In addition to the several formats represented by all these item types, each module (V, M, TSWE) has detailed content specifications, which cover such aspects as point tested, subject matter, item content, length, and so on, and detailed statistical specifications.

Coding Categories

Most earlier studies investigating possible aspects of test bias have tended to look at the few outlying items yielded by subgroup analyses and to attempt to assign causes of aberrant behavior by identifying patterns in the outliers. Most often, outlying items are so few in number that they defy the detection of patterns. In this study and in related studies with the GRE and GMAT (O'Neill, Wild, & McPeek, in progress), the typical procedure was reversed. That is, several hypotheses regarding potential causes of group differences in performance were advanced and the data were examined to see whether or not these hypotheses were confirmed. These hypotheses were examined using an elaborate item coding system based on aspects of test items that might lead to group differences. In the formulation of the categories to be coded, the objective was to include all aspects in which test items that

were ordinarily similar might differ. Thus, for example, although an Antonym always presents a word and asks students to choose the most-nearly opposite word or phrase, Antonyms differ in several respects. Some Antonyms are drawn from the field of the Humanities (e.g., "lyrical"), some from Science (e.g., "arid"), some from the World of Practical Affairs (e.g., "inflation"), and some from Human Relationships (e.g., "anger"). Within each of these content areas, there are further breakdowns (e.g., technical vs. non-technical Science, philosophy vs. art in Humanities). Some Antonyms represent concrete entities (e.g., "anaesthetic"), while some are more abstract (e.g., "obsession"). Some are derived from Anglo-Saxon words, while some are Graeco-Latin in origin and may be cognates of similar words in the Romance languages (e.g., "inundation"). And so on. While most coding categories tend to represent tangible aspects of items, a few were included that are less tangible (e.g., emotional content). In short, all differences and therefore all potential causes of different performance were sought.

In setting up coding categories, the investigators started with all of the test specifications as categories; the test specifications cover some but not all of the many variations in items. Results of other investigations that had suggested hypotheses for differences, as well as categories used in other simultaneous studies (e.g., by GRE and GMAT), led to additional categories. Further, meetings with test specialists in the Verbal, Mathematics, and TSWE areas, in which specialists were asked to specify the possible ways in which test items could differ from each other, led to yet more categories. In all, far more than 100 coding categories were generated. Most of these categories are specific to item type, but several (e.g., subject-matter content) cross

item type. (See appendices A, B, and C for the V, M, and TSWE categories used.)

Each item in each test was then coded for the presence or absence of each attribute represented by a category. Items were double coded independently by ETS professional staff, and differences in the assigning of codes for individual items were resolved through discussion.

Analyses

Item Analyses

On each form of the test, standard item analyses were performed for each of the reference and focal groups. For each group, these procedures yielded test summary information for each of the forms as well as a statistically descriptive accounting of aggregate candidate responses for each item.

Mantel-Haenszel

The Mantel-Haenszel procedure was used to investigate differential item functioning. This procedure, which has been refined and described by Holland and Thayer (1986), is a noniterative contingency table method for detecting test items that function differently in two groups of examinees. The procedure assumes that if test-takers know approximately the same amount about what is being measured, then they should perform in much the same way on an individual test question (or item) regardless of their group membership (e.g., sex, race, ethnicity). For each reference/focal group comparison, the Mantel-Haenszel procedure provides a single summary measure of the magnitude of the differential item functioning (DIF). DIF occurs when examinees from the reference group and the focal group who have comparable overall performance on

the test (or some other relevant matching criterion) evidence markedly different performance or success on a particular item.

DIF is expressed as differences on the delta scale, which is a scale used at ETS to indicate the difficulty of test items. For the DIF statistic, a value of 1.00 means that examinees in one of the two groups being analyzed found the question to be one delta point (about 10%) more difficult than did comparable or matched examinees in the other group. A negative value means that the item is differentially more difficult for the focal group, a positive value that the item is differentially more difficult for the reference group.

In this investigation, the Mantel-Haenszel statistic, or DIF, was computed for each item for each reference/focal group comparison. It allowed us to investigate whether there were unexpected group differences in item functioning that were present after groups had been matched with respect to the ability that was being measured by the test. Thus, students were matched on overall formula score on each of the three scales. That is, total Verbal formula score was used as the matching criterion for the Verbal sections, total Mathematics formula score as the matching criterion for the Mathematics sections, and total TSWE formula score as the matching criterion for the TSWE sections.

DIF is used at various stages of the test development and review process. Initially, when SAT items are pretested, items with extreme DIF values are flagged and reevaluated. At this stage an item with large DIF is revised or eliminated from the item pool. In instances that involve tests with insufficient numbers of focal group members at the pretest stage, DIF is calculated at the stage of test scoring just prior to score reporting. At this point, if an item is removed, the data are purified; that is, analyses

are rerun without the "suspect" item. Finally, DIF is used retrospectively, as in this investigation, in order to test larger groups and in order to generate and test hypotheses that might shed light on why groups comparable in ability performed differently. Ultimately, it is expected that the information gained will lead to changes or improvements in future tests. In this study, the data were not purified; that is, groups were matched using a total test score that may include some items with extreme DIF values.

Analysis of Variance

In addition to making it possible to identify individual items that exhibit unexpectedly differential performance, the Mantel-Haenszel is useful in evaluating whether there are categories of items exhibiting differential performance. In past studies, when items with large DIF were identified, the individual items were scrutinized to determine whether there was a discernible reason that the differences had occurred. Although in such cases it is sometimes possible to hypothesize about why a particular characteristic causes an individual item to have a large DIF value, not every item with that characteristic is associated with a DIF value that supports the hypothesized rationale. Further, every item has a myriad of characteristics (e.g., short or long stem, gender or ethnic reference, presence of homographs, subject-matter content, etc.). Attempting to isolate a specific characteristic based on the few items that are identified with large DIF values in a test form very often becomes more speculative than empirical.

In this investigation, a *priori* categories of item characteristics were identified and items were coded accordingly. One-way analysis of variance techniques were used to identify categories of item characteristics that

resulted in significant differences between the reference and focal groups. The Mantel-Haenszel statistic for each item served as the dependent measure for each reference/focal group comparison. Analyses were performed on each of the six forms individually and on the combined six forms. The combination of forms allowed the number of items in categories that occur relatively less frequently to be aggregated, with the benefit that the results are more reliable and also far less dependent on possible idiosyncrasies in one test form.

Results

Results of the analyses will be presented in a number of different ways, each representing a significant and unique point. First, a set of overall results for each of the reference group/focal group comparisons will be given. These overall results for each comparison will start at the most general level and will become increasingly specific. That is, mean score differences between groups on each of the three SAT scales (Verbal, Mathematics, TSWE) will be presented first. These results reflect differences that exist in mean performance between the reference and focal groups prior to matching.

The remaining results presented in this report refer to differences that emerged once comparison groups had been matched on overall performance. As such, these results cannot be said to reflect absolute differences between groups. For example, it is inappropriate to say on the basis of these results that females perform better (or worse) than males on items with a particular characteristic. Rather, significant findings indicate that when males and females have been matched on overall performance, females perform relatively better (or worse) than males on items with a particular characteristic as

compared to their relative performance on items without this characteristic. In some instances, this means that the difference or disparity in performance (favoring one group or the other) is significantly greater on some items than on others. In other instances, items with a characteristic favor one group whereas items without the characteristic favor the other group. The analyses of variance focus on unexpected differential performance between the matched focal and reference group rather than on differences in actual levels of item difficulty.

Initially, results will be provided that deal with summary information regarding DIF values. Average DIF values for each group for each item type on each scale are provided. Then the number and percent of items with relatively large DIF values--again, for each group on each item type on each scale--are presented. Following these overall results will be a presentation of results of the analyses of variance for each scale and within each scale for three specific item coding subcategories: (1) points tested, (2) subject matter covered, and (3) item format. Although the three scales will be handled separately (since the matching criterion for students was done separately by scale), where similarities or obvious differences among scales exist, these will be pointed out.

Overall Performance

Table 2 presents a summary of mean formula score differences between

---

Insert Table 2 about here

---

reference and focal groups in standard deviation units. This table is

presented for overall reference only; its results present differences before matching and therefore represent absolute impact. On the Verbal scale, for all six forms, focal groups, with the exception of Asian-Americans, performed less well than reference groups. On the six forms, the male/female difference ranged from -.11 to -.15 of a standard deviation on overall Verbal performance. White/Black differences averaged about one standard deviation, while White/Hispanic differences averaged about two-thirds of a standard deviation. For the White/Asian-American comparison, the picture is somewhat different and less consistent. On two forms, Asian-American differences from Whites were negligible (.02 and -.02), while performance on the other four forms ranged from .07 to -.16.

On the Mathematics Scale, male/female differences were substantially greater than those found on the Verbal scale. Females ranged--again, consistently--from -.38 to -.46 of a standard deviation lower than males. For the White/Black comparison, differences in Mathematics were about the same as differences in Verbal: about one standard deviation. White/Hispanic differences were about the same as -- or very slightly less than -- Verbal differences: from -.49 to -.63 of a standard deviation. Only for the White/Asian-American comparison in Mathematics did the focal group fare consistently better: Asian-Americans performed about one-third of a standard deviation better than Whites.

On the TSWE scale, females outperformed males by approximately .12 standard deviation, with a range across the six forms of .06 to .17. Whites performed better than Asian-Americans by about one-quarter of a standard deviation (range of -.20 to -.42); Whites performed better than Hispanics across the six TSWE forms by an average of about one-half of a standard

deviation. White/Black differences averaged slightly less than one standard deviation, with a range of -.85 to -1.02.

Table 3 presents average DIF values for reference and focal groups for each item type in each scale; negative values indicate relatively poorer performance by the focal group, while positive values indicate better performance. Since values in this table represent differences after groups

---

Insert Table 3 about here

---

have been matched for performance on the SAT overall scales, the values indicate not absolute impact but rather the relative ease or difficulty of each item type on each scale for each reference and focal group. On the Verbal scale, females fared better than matched males in Reading Comprehension. Antonyms and Analogies tied for the Verbal item type on which females fared, on average, least well compared to matched males. This finding, suggesting that females seem to do relatively better than matched males on item types with more context and relatively poorer than matched males on those with little or no context, is consistent with earlier findings on the SAT (Wendler & Carlton, 1987).

The pattern across ethnic groups within the Verbal scale is somewhat inconsistent, with the possible exception of Analogies, on which all groups performed less well relative to matched groups of Whites. The mean disparity on Analogies was smallest for Asian-Americans/Whites and, in fact, mean Asian-American/White DIF values were relatively small across all Verbal item types. Review of mean Black/White DIF values suggested that, on average, Whites performed better than matched Blacks on Analogies whereas the reverse was true

for Antonyms. Hispanics performed less well than matched Whites on Analogies and Sentence Completions and better than matched Whites on Reading Comprehension.

The Mathematics scale showed very little difference in performance between its two item types. Differences in the item types across the board between matched focal and reference groups were negligible.

On the TSWE scale, the reference/focal group performance disparity was greater on Sentence Correction (which requires a choice of the best written sentence) than on Usage (which requires a choice of if and where in a sentence an error occurs), with all focal groups performing, on the average, relatively better than matched reference groups on Sentence Correction. Differences here, however, were large enough to be significant only for the male/female and White/Asian-American comparisons. Here, as in the Verbal scale, one might speculate that females and minority groups do unexpectedly better when the task provides more context and less exact pinpointing. Other content and stylistic item type aspects that contribute to differential functioning will be discussed later in this section.

Tables 4 and 5 present the number and percent of items in

Insert Tables 4 and 5 about here

---

each scale and each item type that evidenced relatively extreme DIF Values. As in the previous table, values represent differences after matching and as such are valuable for detecting, within scales and within item types, the existence and the magnitude of performance differences between groups matched on overall scale score. The previous table dealt with all of the test items

within an item type; these two deal only with items that show large negative or positive DIF. Table 4 deals with items with DIF values greater than 1.0 or less than -1.0. The greatest percentages of negative DIF items on the Verbal scale tended to be Analogies and Antonyms for all focal groups. Differences in positive DIF items both were smaller and showed no consistent pattern. On the Mathematics scale, differences between the two item types for matched reference/focal groups were negligible. In TSWE, the number of extreme DIF items for the male/female comparison (two negative, one positive) was negligible. For the matched White/ethnic group comparisons, the number was quite low for Black and Hispanic comparisons and appreciably higher (both negative and positive) for the White/Asian-American comparisons.

Table 5 presents results of test items with DIF values greater than 1.5 or less than -1.5. At these greater values of DIF, the number (and proportion) of items fell precipitously--by an average factor of more than three--for all focal groups in all three scales and for all item types within each scale. Patterns, however, remained the same. With the exception of the White/Asian-American comparison, negative and positive DIF values were greatest in the Verbal scale for all matched comparison groups; within the Verbal scale, Analogies and Antonyms tended to exhibit the largest number of extreme DIF items, Sentence Completions and Reading Comprehension the least. The number of extreme DIF items in both Mathematics and TSWE was so negligible as to almost preclude interpretation. Regular Mathematics was somewhat more likely than Quantitative Comparisons to produce extreme DIF items; as before, in TSWE, Usage had more large DIF items for White/ethnic matched groups than did Sentence Correction. There were no items in the male/female comparison in either item type of TSWE with DIF values greater than 1.5, or less than -1.5.

Item Category Results

The foregoing results dealt with differences at a macro-level, that is, at the level of the item type only, regardless of its subject-matter variations and format variations. The remaining discussion will deal with the "smaller" aspects of items, seeking to identify those specific characteristics that most divide performance among groups of test candidates matched for overall ability. These aspects relate to both the content and the format of items.

In the sections that follow, summary item category data will be presented in tables and interpreted in the text. For consistency, all tables will indicate whether an obtained F statistic was significant at the .10, .05, or .01 level of significance. The authors recognize that these are very liberal designations, particularly given the large number of analyses of variance that were performed. Undoubtedly, there will be errors in the overidentification of group differences. However, a major purpose of this investigation was exploratory, and one objective was to look for possible patterns of performance. Patterns that emerge across reference/focal group comparisons may suggest that there is a confounding of the item category with some other item characteristic, or they may suggest that there is something about the item category favoring Whites and/or males that is associated with such factors as test anxiety or expectations or some element associated with the White, majority culture. In either case, further review is warranted.

In the interpretation of results, extreme caution should be used when drawing conclusions about group differences in performance. It is critical to keep in mind that each reference/focal group comparison was made on groups matched on overall performance on the relevant SAT scale. In addition, a

statistically significant F value indicates only that there were unexpected reference/focal group differences in the mean Mantel-Haenszel DIF values (which represent relative differences in item functioning) for at least two of the item categories in the analysis. In many of the item categories there were too few items for the results to be considered reliable. When this occurred, it has been noted in the text. Further, it must be kept in mind that isolated findings may be spurious and only suggestive of an area for further consideration.

Points Tested

With regard to content, skills tests such as the SAT have two major divisions. First, there is the area of points tested (e.g., vocabulary knowledge, tense sequence, percentage), and second, there is the area of the subject matter that forms the surrounding matter of items. To illustrate: one can test tense sequence in a sentence dealing with science ("The test tubes were broken before she started the experiment") or in one dealing with literature ("Milton wrote 'Paradise Lost' before he wrote 'Paradise Regained'"). In the section that follows, the focus is on points tested independent of the content in which it is embedded.

Verbal Scale: On the Verbal scale, the subject matter is far-ranging and, moreover, is not typically taught in courses or course sequences: high school courses dealing with Antonyms, Analogies, Sentence Completions, and even Reading Comprehension are largely, if not entirely, nonexistent. As a result, the Points Tested on the Verbal scale are quite few in number and,

further, the few that were examined showed very few group differences in performance.

The one variable examined here with Analogies was Semantic Relationship, that is, the relationship between the first and second terms in an Analogy. Table 6 provides mean DIF values for the item categories. What differences

---

Insert Table 6 about here

---

exist were so small as to be nonsignificant in the male/female comparison and White/Asian-American comparison. Differences in the White/Black comparison and the White/Hispanic comparison (and, incidentally, mimicked somewhat in the White/Asian-American comparison though not to a significant degree) indicated that these focal groups did less well relative to matched groups of Whites on Part-Whole relationships (e.g., "tree is to forest as....") and relatively better than matched groups of Whites on Contrast relationships (e.g., "miser is to generous as...."). It is difficult to interpret the meaning of these results, but it is worth pointing out that the number of test items in each category here was so small as to call the results into question regarding both practical and statistical significance.

With Antonyms, the findings were somewhat similar: each of the two variables examined was significant for only one comparison and results are difficult to interpret. (See Table 7.)

Insert Table 7 about here

The first finding was that Hispanics seemed to do relatively better than

matched Whites on Antonym items in which a Fine Distinction, rather than a General Distinction, was required among answer choices. For the second variable, Parts of Speech, Whites did relatively better than matched Asian-Americans when the Antonym was a Verb as compared to their relative performance when the Antonym was a Noun.

In Reading Comprehension, the only variable examined in Points Tested was the kind of reading question asked. (See Table 8). No consistent pattern emerged here, except perhaps that minority groups tended to fare relatively better, when compared to matched groups of Whites, on the more global

---

Insert Table 8 about here

---

questions (i.e., questions on the meaning of the whole passage) than their relative performance on questions regarding analysis of structure, logic, and style. Again, the numbers of items in most categories here tended to be small.

Mathematics Scale: For Mathematics, the situation was somewhat different, since there was a more clearly definable body of knowledge whose mastery or nonmastery could be readily described. Consequently, for Points Tested in Mathematics, several hypotheses were tested. Table 9 presents results for the more global categories of Points Tested. As was discussed

---

Insert Table 9 about here

---

earlier, the slight difference between the Quantitative Comparison item type

and the Regular, or problem-solving, Mathematics item type was not significant. With regard to the Primary Content of mathematics items, women did relatively better than matched males on items categorized as Miscellaneous (particularly Number Sets and Number Systems), whereas they performed relatively less well, when compared to matched males, on Geometry. For both Asian-Americans and Hispanics, when compared to matched groups of Whites, this finding was reversed: Asian-Americans and Hispanics performed relatively better than matched groups of Whites on Geometry, particularly as compared to their performance relative to matched Whites on the items classified as Miscellaneous.

On the related Multiple Categories variable, females performed better than matched males when Arithmetic/Algebra was required, whereas the reverse was true when Arithmetic/Geometry was required. These findings are consistent with those of Doolittle and Cleary (1987). Blacks, relative to matched Whites, performed relatively better when Arithmetic/Algebra was required as compared to their relative performance when Arithmetic alone was required. Again consistent with the earlier finding, Asian-Americans and Hispanics did relatively better than matched Whites on categories that included Geometry (i.e., Arithmetic/Geometry and Algebra/Geometry) as compared to Arithmetic Only, which was associated with weaker performance by these focal groups relative to matched groups of Whites.

A related finding deals with whether or not the item contains a variable (e.g., an unknown, such as x, y, etc). All focal groups, when compared to matched reference groups, performed relatively better when a variable was present than when it was not present, whereas reference groups performed

relatively better when there was no variable in the item or in the stimulus that accompanied one item.

An interesting finding was performance on the variable Ability Level, which progresses in cognitive complexity. With the exception of Factual Knowledge, based on only one item and therefore not considered, the other cognitive levels seemed to show--for females and Asian-Americans--a steady shift in relative performance compared to the reference groups, from Mathematics Manipulation through Higher Mental [Processes]. Average DIF values suggested relatively stronger focal group performance on mathematics items requiring lower level mental processing, as compared to mathematics items requiring higher mental processes, which seemed to be associated with relatively stronger reference group performance. Findings for Hispanics were consistent with this trend but were not significant.

When the variable was Type of Solution, the average discrepancy between reference/focal group performance was greater, with female, Black, and Hispanic groups consistently performing relatively better than their matched reference groups on items in which the solution was General as opposed to actually Computed. White/Asian-American results were consistent with this pattern but not significant. Whether or not this related to a relative lack of precision (as perhaps seen in the tendency for a relatively poorer focal group performance in Arithmetic and Arithmetic Only in two previous discussed categories) is open to conjecture.

The final category in this table is Special Topics, broken down so finely that interpretations should be made warily. Of note, however, was the fact that on items involving Angles and Linear Measure, while females did relatively less well than matched males, Hispanics and Asian-Americans did

relatively better than matched groups of Whites, as was the case with the related Geometry discussed earlier.

Table 10 presents the results of Points Tested in breakdowns within

---

Insert Table 10 about here

---

the primary content areas of Arithmetic, Algebra, Geometry, and Miscellaneous. Of most note here were several consistent patterns within Algebra: the relative strength of all focal groups in all Algebraic Operations as compared to the relative weakness of all the focal groups in Word Problems.
Table 11 presents DIF information on a series of spatial/visual factors.

Insert Table 11 about here

Of possible note here was the finding that the discrepancy between male/female performance, favoring males, tended to be greater on items with a spatial component as compared to relative performance when no spatial component was involved. Specifically, when Figures, Graphs, or Tables were present, women performed relatively less well than when there was no such stimulus.
Similarly, Blacks found items with Figures more difficult than did matched Whites, as compared to items where there was no Stimulus Format or Figure, on which Blacks performed relatively better than matched Whites.

On the other hand, Asian-Americans and Hispanics seemed to do relatively better than matched groups of Whites on items with a spatial component. These two focal groups performed relatively better on items involving Geometry and Estimation and also relatively better with items containing Figures, as

compared to their relative performance on the other item categories within these variables. Items involving Figures (particularly when the figure was not provided) consistently were associated with a performance discrepancy favoring Asian-Americans and Hispanics, as compared with relative performance on items not involving Figures.

TSWE Scale: Table 12 presents differential performance as a function of Points Tested in TSWE. Here, as in Mathematics, there is a body of relatively finite knowledge, that is, points of grammar and usage. Points tested fall into the two item-type dichotomy discussed earlier and into the elements of

---

Insert Table 12 about here

---

grammar and usage enumerated in the table under the heading of Specification. Differences by point tested were highly significant for all groups. The findings suggested that the area of Subject/Verb Agreement was an area of relative difficulty for all focal groups compared to their respective matched groups, whereas detecting unwarranted Shifts in sentences and the related detecting of lack of Parallelism were associated with relatively stronger performance by all focal groups compared to matched reference groups. Again, as in the Verbal Scale, one might conjecture that focal groups fared better on the larger elements (as seen in Shift and Parallelism) than in the smaller elements (as represented by Subject/Verb Agreement).

Item Content

The second major aspect of content (in addition to Points Tested) is kind of language (e.g., Technical) and subject matter (e.g., Science, Human

Relations, etc.), that is, the actual subject matter and words used--although not tested for--in writing sentences and passages and in framing questions. On all three scales, it was in this area that some of the largest sets of differences were found between the matched reference and focal groups.

Subject Matter: On the Verbal scale, all four item types were divided into various subject matter disciplines--in the test specifications as well as in this study. For the discrete item types--that is, Antonyms, Analogies, and Sentence Completions--these were Aesthetics/Philosophy, the World of Practical Affairs (e.g., wars, politics, sports, business), Science, and Human Relationships (e.g., emotions, everyday interactions). It is worth repeating that knowledge of these subject matter disciplines per se is not tested for in the SAT. Rather, on each SAT form, Verbal items are distributed evenly across these areas in order to ensure that students are neither systematically advantaged nor systematically disadvantaged by degree of previous exposure to or familiarity or comfort with any one field. Table 13 presents the results of the content breakdowns with regard to the subject matter of the Verbal discrete item types. In each case, a Mixed/Overlapping Category was added for items that either contained more than one subject area or were on the border

---

Insert Table 13 about here

---

between two areas. In addition, since Science had in earlier studies been shown to be associated with differential item performance, the Science category for Analogies was broken down into several subareas in an attempt to analyze which kinds of Science were associated with differential performance.

Data for these specific fields of Science are provided; the numbers of items in each area, however, were too small to permit the drawing of conclusions.

For the subject matter item content category, results for females on all three discrete item types were significant and highly consistent. Results suggested performance discrepancies favoring females on items in the fields of Aesthetics and Human Relationships and performance discrepancies favoring males on items in Science and in the World of Practical Affairs. This is consistent with findings on Analogies in the General Test of the Graduate Record Examinations (Pearlman, 1987).

For reference/focal group comparisons involving ethnic groups, six of the nine subject matter analyses resulted in significant findings. In each significant case, Science was associated with a performance discrepancy favoring the White reference groups whereas Human Relations was associated with relatively better performance by the focal groups. This relationship was found for Blacks on all three discrete item types, for Hispanics on Antonyms and Analogies, and for Asian-Americans on Analogies.

Table 14 presents results for Reading Comprehension. Again, results

---

Insert Table 14 about here

---

indicated that females performed relatively better in the field of Humanities and on Narrative passages (a breakdown of Humanities) than did matched males, whereas males again performed relatively better on the items with Science content. Females also performed relatively better than matched males in Social Science. These findings held for the expository Humanities and Social Science passages as well as for the Argumentative--or Persuasive--ones. Also,

females performed less well relative to a matched group of males on items based on Technical Science passages, particularly as compared to their relatively stronger (compared to matched males) performance on other, non-Science Reading Comprehension items.

For ethnic groups, the trend apparent in the discrete Verbal item types did not hold consistently in Reading Comprehension. No significant results were obtained for Blacks or Hispanics. In Reading Comprehension, Asian-Americans, unlike females and unlike their own performance on Analogies, tended to perform relatively better than the matched reference group on Science passages, particularly Technical ones, as compared to their relative performance on Humanities passages, whether Literature, Narrative, or Other, and whether Expository or Argumentative. This disjunction in the Asian-American group is hard to explain. One may conjecture, however, that possible language differences for Asian-Americans (even those who reported English to be their best language), compared to matched Whites, were either exacerbated in the more literary or general passages and/or modified by technical language, which might be more familiar and more accessible to students from an Asian-language background.

Moving to TSWE--where items are the smaller units of sentences rather than passages--we see in Table 15 some of the same results for focal groups as

---

Insert Table 15 about here

---

with the Verbal discretes. For Asian-Americans, Hispanics, and females, compared to matched reference groups, there were performance discrepancies favoring the reference groups on sentences with Science content. This was

contrasted with a relatively stronger performance (compared to matched reference groups) by Asian-Americans and Hispanics on Social Science sentences and by females and Hispanics on sentences dealing with Student Relevant concerns and Everyday Activities. Results for Blacks on TSWE Subject Matter breakdowns, while not significant, were nonetheless interesting in that they supported the tendency for focal groups to experience relatively more difficulty with items with a Science context.

Table 16 presents the results of two kinds of content breakdowns on the Mathematics scale. Results for both breakdowns and for all focal groups were

---

Insert Table 16 about here

---

remarkably consistent and highly significant. Females, Asian-Americans, Blacks, and Hispanics performed relatively better than matched reference groups when Mathematics problems were Abstract (unapplied or not drawn from "real life"), whereas the reverse was true when Mathematics items were in the form of "real life" or Word Problems. The other--and related--category, Relation to Curriculum, showed similar results. Performance for females and for all ethnic groups was relatively better than for matched reference groups on mathematics items that were very much like problems in mathematics textbooks as compared to when the items departed from textbook-like characteristics. Relative to their matched reference group peers, then, focal groups members did relatively better when dealing with "pure" mathematical manipulation and relatively worse compared to matched reference group members when asked to extrapolate or to apply what they had learned. The extent to which this may be experientially caused for females or linguistically related

for at least the Asian-American and Hispanic focal groups is conjectural; further research would seem called for here, given the magnitude and consistency of the findings.

Technical/Non-Technical: Tables 17, 18, 19, and 20 present the results of analyzing the Verbal scale for other and related content variables.

Insert Tables 17, 18, 19, and 20 about here

The first, Table 17, deals with the extent to which Verbal items are Technical or Non-Technical and the relationship of this variable to differential item performance. For Sentence Completions and Antonyms, the limited number of test items distributed in many of the cells precluded interpretation. Significant results were obtained only for Reading Comprehension and for Analogies. In Reading Comprehension, females performed relatively better than matched males on test items for which both the passages and the questions were Non-Technical as compared to relatively worse performance when both components were Technical. On Analogies, similar results were obtained: females performed just slightly better than matched males on Non-Technical items and relatively less well than the males on items that had been classified as Technical. These findings are in accord with the Science/Non-Science breakdowns mentioned earlier.

Also as before, Asian-Americans reversed this trend, performing relatively better than matched Whites when Reading Comprehension passages and questions were both Technical and relatively less well when they were Non-Technical or when only the Passage was Technical. For Hispanics, there was a

marginally significant tendency for the White/Hispanic discrepancy to be slightly larger when Reading Comprehension questions and passages were Technical than when only the passages were Technical. Results for Blacks were nonsignificant. On Analogies, results were nonsignificant for Asian-Americans and Hispanics and inconsistent and difficult to interpret for Blacks.

Concrete/Abstract: Drawing on the suggestion in the Wendler and Carlton (1987) study that female performance might be enhanced on items in which the terms were Abstract and relatively disadvantaged on items in which the terms were Concrete, the investigators examined Analogies on the Concrete-Abstract continuum. Since coding of the other Verbal item types resulted in insufficient items in each cell, only Analogies were considered for this variable. Table 18 indicates that, as suspected, there was a trend for females to do relatively better than matched males on Abstract Analogies and relatively worse than matched males on those Analogies containing all Concrete terms. Supporting the results of Freedle, Kostin, and Schwartz (1987), the same pattern of relatively stronger performance with Abstract terms and relatively poorer performance with Concrete terms was found for all of the ethnic groups, and with striking consistency. This startling finding leads to the question of whether the relatively better performance of focal groups in the areas of Humanities and Human Relations compared to their relatively poorer performance in the areas of Science and Practical Affairs is really completely a function of "discomfort" with or lack of exposure to the latter two fields or of whether these two fields tend to contain more Concrete than Abstract terms. That is, to what extent is performance in subject matter areas dependent on the Abstract vs. Concrete nature of the terms used? Clearly,

future research in which the subject matter and the concreteness of the terms are unconfounded would seem warranted.

Latinate Language: Table 19 presents results of coding items for the absence or presence of Graeco/Latin word origins. For Analogies, the significant results for Asian-Americans, Blacks, and Hispanics indicated that focal groups performed relatively better than matched Whites on Latinate items and relatively poorer than matched Whites on items containing words without Classical origins. For Hispanics, this lends support to findings by Schmitt (1988) and is consistent with the expectation that those familiar with Spanish would perform better on words that represent cognates or near-cognates, as Latinate words do. For Asian-Americans, this finding might be associated with the result found earlier in which Asian-Americans performed relatively better on items with Technical language (hence, Latinate). For Blacks, the relatively better performance with Latinate words is clearly present but difficult to interpret. The Latinate variable yielded no significant results for Antonyms.

Homographs: Table 20 presents results when Analogies and Antonyms were coded for the absence or presence of Homographs (words spelled the same but having different meanings, as in "bear," "bark," "press," etc.). This item category was considered by Schmitt (1988) in a study of Hispanic examinees; however, there were too few items with homographs in that study to judge their impact. For all ethnic groups on Analogies, the presence of Homographs tended to be associated with relatively poorer performance compared to matched groups of Whites. It is possible that ethnic group performance may have been more

disrupted than matched White performance by the potentially confusing appearance of words that looked like but were different from more commonly appearing words. Results on this variable for Antonyms were nonsignificant.

Parts of Speech: Table 21 presents the results of looking at Parts of Speech for the words in the stems and keys of Analogies and Antonyms and in the keys of Sentence Completions. No consistent patterns were found to be associated with Parts of Speech, and only four analyses yielded significant

---

Insert Table 21 about here

---

results. Two of these were that, with Analogies, Blacks and Hispanics, relative to matched Whites, performed less well when the two terms were both Nouns or both Verbs as compared to when the two terms were mixed parts of speech or both Adjectives. Females and Asian-Americans followed the same trend, but results were not significant. An explanation of this curious pattern is not obvious, although it should be noted that the numbers of items with both Verbs or both Adjectives were too small for reliability. One possibility is that Nouns tended to be concrete, and, as was noted earlier, focal groups tended to perform less well on concrete Analogies than on abstract Analogies. A third significant finding was that with Antonyms, Asian-Americans, when compared to matched Whites, performed relatively better with Nouns than with Verbs, a finding paralleled, though nonsignificantly, by females and Hispanics but not by Blacks. Finally, the finding that with Sentence Completions, Blacks tended to do relatively better with two Nouns or one Noun in the key position is difficult to explain.

Emotive or Controversial Material: Tables 22 and 23 present results for the Verbal and TSWE scales on a group of related variables having to do with the degree of Emotion or Controversy in the items. This analysis was

---

Insert Tables 22 and 23 about here

---

attempted since an earlier study (Wendler and Carlton, 1987) suggested that females might do comparatively worse with material that is upsetting or controversial. This was a difficult variable to work with since test development guidelines dictate that all potentially upsetting material be eliminated from tests. What is left, then, is relatively mild or neutral material, with very few items likely to be coded Strong in the Emotive category. This was evident from the results presented in Table 22. For the most part, results were not significant. On Sentence Completions, there was a tendency for females to perform less well relative to a matched group of males on the few items with Strong Emotive Language as compared to their relative performance on more Neutral items. Table 23 presents findings with the related variables of Controversial Language for discrete Verbal items and TSWE items and for Socially Relevant (and hence potentially controversial) Sentence Completions and passages for Reading Comprehension. Females consistently and significantly performed relatively worse than matched males on the Controversial items and on the Socially Relevant Sentence Completions as compared to their relative performance on Sentence Completions and Analogies without these characteristics. Only the results for Reading Comprehension items were not consistent with this trend, and these findings were not significant. Taken together, the findings from these related variables lend

tentative support to the Wendler and Carlton hypothesis: Females may be more readily disrupted on some Verbal item types when the content of the item is relatively Emotive or Controversial.

The impact of Emotive or Controversial item content on ethnic group performance was not consistent. Results were largely not significant. Hispanics, relative to matched Whites, tended to show a trend similar to that of females on Analogies and Antonyms with Strong Emotive content. The reverse was true for Hispanics on Controversial Analogies. Similarly, there was a tendency for Blacks to perform relatively better than matched Whites on Controversial Analogies as compared to their relative performance on Analogies with Neutral material.

Minority and Gender Reference: The final set of content variables relates to Minority and Gender References and to references to People in general, particularly in passages and sentences on the Verbal and TSWE Scales. Table 24 presents results with the variable Minority Stimulus in Sentence Completions and Reading Comprehension, and Table 25 presents results for Gender Reference in the same two item types. Although the coding system

---

Insert Tables 24 and 25 about here

---

allowed for coding for all ethnic minorities, the six forms analyzed actually named only those specified in the tables; in many instances, the numbers of items in a cell were too small to be considered reliable. In both item types, females performed relatively better than matched males when People were referred to as compared to when there was no reference to People, even when

ethnic group or Gender was not identifiable; in Sentence Completions, females did relatively better when People or ethnic group or Gender were specifically named. In general, who was named or referred to did not seem to matter much; what did matter in female performance was the presence or absence of People.

In Sentence Completions, no ethnic group had significant results for references to or naming of ethnic Minorities. In Reading Comprehension, Blacks performed relatively better than matched Whites in passages that referred to or named Black Americans, as compared to when no one was named or referred to. Hispanics also appeared to do relatively better than matched Whites when Minorities were named than when no one was named, and there was a slight tendency for Asian-Americans (but not Hispanics) to do relatively better when Hispanics were referred to. The only general pattern, albeit weak, was that the reference to or naming of ethnic group members tended to lead to relatively better performance by focal group members. Ethnic groups, when compared to matched Whites, also tended to perform relatively better when Gender references were made than when there were no such references.

TSWE shows similar patterns for females for Gender Reference, as seen in Table 26. Females performed relatively better than matched males when People, especially females, were referred to in TSWE items than when no People appeared in the items. Hispanics performed relatively better than matched Whites on the items with Female references as compared to their relative performance on items with Mixed or No Gender references. Results for the

---

Insert Table 26 about here

---

other focal groups were not significant. With regard to Minority Stimulus in

TSWE items, no significant differences appeared for reference/ethnic group comparisons, and the values for the male/female comparisons were difficult to interpret, principally because of the few numbers of items in each cell.

Since People referred to in Analogies and Antonyms are not identifiable by Gender or Minority group, these analyses could not be run, but results for references to any People in these two item types are presented in Table 27. There was a consistent pattern for Asian-Americans, Blacks, and Hispanics

---

Insert Table 27 about here

(but not females) to perform relatively better than matched reference groups when People were referred to in Analogies than when they were not. No significant differences emerged with Antonyms. Also, the number of Antonym items with People References was very low.

A final table in this set, Table 28, presents results for Gender Reference and Minority Stimulus on the Mathematics scale. As is evident, only one item made

---

Insert Table 28 about here

---

specific reference to a Minority, and references to Males or Females were limited. Consequently, it was more useful to focus on the presence or absence of People. There was a consistent pattern for all focal groups to perform relatively worse than matched reference groups when People were referred to as compared to when no People were present. At first glance, this startlingly consistent and significant finding seems at odds with the findings in the Verbal and TSWE scales, where focal groups tended to perform relatively better when People were referred to. What the results

with Mathematics most likely represent, however, is not a contradiction but rather a confirmation of the earlier finding in Mathematics that focal groups tended to perform less well relative to matched reference groups with word problems or real-world problems than with abstract or textbook-like Mathematics problems. Put simply, Mathematics items that refer to People (as those represented in Table 28 do) were likely to be relatively harder for focal groups because they departed from the "pure mathematics" problems typically found in textbooks. As such, they may have disrupted rather than enhanced performance for females and for ethnic group members (compared to their matched counterparts).

Item Format

A final group of variables that was studied related to the format of test items, or the formal characteristics of test items. Based on obvious format differences and on previous studies, many variables were examined. Of these, a few were found that differentiated focal group performance from reference group performance, while others showed no consistent patterns of difference.

Length: Tables 29 and 30 present a summary of the analyses of variance results of an examination of several aspects of Length in items on all three scales.

Insert Tables 29 and 30 about here

In both of these tables, results are presented in abbreviated format (i.e., without means and standard deviations and with levels of significance only where appropriate.) Table 29, dealing with the Verbal scale and TSWE, indicates very few significant findings (only slightly more than would be expected by chance). In

Sentence Completions, increased Length of Stem was associated with relatively better performance by Asian-Americans relative to matched Whites; in Reading Comprehension, Length of the Stimulus (or passage) was associated marginally with relatively better performance by Hispanics compared to matched Whites; and in the Usage item type in TSWE, shorter options were marginally associated with relatively better performance by females. Other Length variables either were nonsignificant or led to inconsistent findings (i.e., as the Length of the Stem or Options increased, there was no consistent pattern of differences in reference/focal group performance).

On the Mathematics scale, Length variables were associated with a more pronounced and slightly more consistent effect. In Regular Mathematics items, with one exception, a long Stem systematically seemed to be associated with relatively poorer performance by all focal groups compared to the matched reference groups. In Quantitative Comparison items, both Length of Stimulus (including charts, graphs, etc.) and Length of Stem significantly were associated with relatively poorer performance by females and Asian-Americans compared to their respective reference groups. It is worth noting here that the added Length in Mathematics items often resulted from items being cast as word problems rather than as straightforward mathematics problems. The relatively poorer performance of focal groups on mathematics word problems, discussed earlier in this section, is most likely what was seen with the Length variables here. That is, it is likely that it was not Length per se that caused problems for the focal groups but rather the kinds of mathematics problems (i.e., word problems) that resulted in longer Stems.

Vertical/Wraparound Response Format: Table 31 provides results for another format variable, whether response options were presented Vertically or whether they were presented Horizontally, in "Wraparound" fashion. Results for Analogies were

striking. All ethnic groups performed less well compared to matched Whites when item options appeared in a Wraparound format as compared to their relative

---

Insert Table 31 about here

---

performance on items with options presented in a Vertical format, on which they performed relatively better than matched Whites; this finding has been replicated using similar methodology for a White/Black comparison on the Graduate Record Examinations General Test (O'Neill, Wild, & McPeek, in progress). Although results for male/female comparisons were not statistically significant, they were consistent with those of the other focal groups. One preliminary hypothesis that could be considered in future research is that this relationship may be associated with some of the past research on field dependence/independence. The Wraparound format might cause more of a distraction for relatively field dependent examinees as they attempt to match the first term in each Analogy option with the first term in the item stem and the second term of each Analogy option with the second term in the stem. When the same format variable was examined for all other item types that lent themselves to this breakdown, no significant consistent effects were noted. It would seem, then, that this variable was associated with significant differences only on Analogies, where a Vertical format appeared to be linked to relatively better focal group performance.

First Appearance: Tables 32 and 33 present results when the First Appearance of an item type (or its subsequent First Reappearance) was compared to "other items," that is, items that follow the first one in a test section. Table 32 gives Verbal and TSWE results. Whether First Appearances and Reappearances in each

section were broken out or combined, results were consistent. First Appearances, as compared to other items on the SAT-V, were associated with unexpectedly poorer

---

Insert Tables 32 and 33 about here

---

performance by ethnic groups relative to matched Whites. Results for females, while not statistically significant, were consistent with this pattern. One possible explanation for these findings is that there could be a "jolt by the unfamiliar" effect, which disrupted focal group performance more than reference group performance. Possibly related findings occurred in the Wheeler and Harris (1981) study, in which females omitted (and therefore got "wrong") first items in the ATP Physics Achievement Test to a much greater degree than did males. Another possible explanation is that the first items in a section may be qualitatively or psychometrically different from other items. However, the First Appearance results seen with Verbal item types were not replicated in TSWE. Nor were they replicated in Mathematics, as seen in Table 33. Reasons for this difference are hard to come by, except for the possibility that the format of TSWE and some Mathematics items could be more familiar (in terms of classroom work)--and therefore less jolting--than the format of some Verbal item types.

Clues to Answer: Tables 34 and 35 present the results of a set of format variables related to where in test items clues to the answers occur. The first three, in Table 34, refer to whether or not the terms in the stem (or question part) of an Analogy come from the same domain as--or can be associated with a word or words in--the options (and this association is independent of the analogical relationship between the two words in the stem), in which case there is a Vertical

relationship and the item is said to be Overlapping. Results indicated that when the key was related Vertically, as compared with when there was no Vertical

---

Insert Tables 34 and 35 about here

---

relationship, the performance of ethnic groups was relatively weaker than that of matched Whites. The same trend was seen for females but it was not significant. When any distractor had a Vertical relationship to the stem, males performed relatively better than matched females, as compared to items without a Vertical relationship, on which females performed relatively better than matched males. Similarly, ethnic group members performed relatively worse than matched Whites when any distractor had a Vertical relationship to the stem, as compared to other items, though this pattern was not significant for Asian-Americans. The relatively poorer performance of ethnic groups compared to matched Whites on the Overlapping category of the Independent/Overlapping variable simply confirms the foregoing, since Overlapping indicates that one or both terms in the stem have either a class/subclass or a subclass/class relationship to one or both of the terms in the key. In all three variables here, relatively poorer performance may have been associated with possible confusion caused by options that were closely related to the stem in subject matter. The tendency for some focal groups to be affected by helpful or deceptive "clues to the answer" is consistent with the findings of Schmitt and Bleistein (1987) for Blacks and Schmitt and Dorans (1987) for Hispanics, in which a verbal or word associative strategy seemed to be more consistently used by ethnic examinees than by matched Whites.

Other Format Variables--Verbal: With very few exceptions, the set of Verbal Item Format variables whose results appear in Table 35 had little effect on performance. The first two refer to when and how Answers are arrived at in Sentence Completions and Reading Comprehension; no significant differences were found here. The third, relating to whether and where Line References appear in Reading Comprehension questions, also yielded no real differences, except perhaps for the finding that the discrepancy between Asian-American and matched White performance appeared to be greater (with Whites performing better than Asian-Americans) when a Specific Line was cited as compared to when Lines were not Referenced; interpretation here is elusive. The final variable deals with the Specificity of stems and options. Females performed relatively better than matched males when the stem was very Specific as compared to when neither stem nor options were Specific. With the ethnic groups, findings were neither significant nor consistent. In general, these variables seemed to have played little part in differentiating focal group performance from reference group performance.

Other Format Variables--Mathematics: Table 36 presents results for several aspects of Item Format in Mathematics. Females and ethnic groups consistently performed relatively better than matched males and Whites on Mathematics items in which Reading Difficulty was coded as Easy as compared to when the reading was coded as Difficult (measured largely by length of stem). Also significant were most findings on the related category of Reading Level (measured largely by complexity of the sentence structure in the stem). Here, all ethnic groups performed less well

---

Insert Table 36 about here

---

relative to matched Whites when Reading Level was judged Difficult as compared to when it was judged Medium or Easy.

The next several categories deal with Item Formats that describe some aspect of the relationship between the stem and the options. The presence of Cannot be Determined and Must/Could attributes could be related to a degree of tentativeness or confidence the examinee feels about his or her responses. These variables, and the Maximum/Minimum Value and Role of Options variables, called upon the examinees to evaluate not only the stem but also the options in selecting their responses. Results for the Maximum/Minimum Value variable were significant across ethnic groups: for Asian-Americans, Blacks, and Hispanics, when the attribute was present, the items were more difficult for the focal groups relative to matched Whites as compared to when the attribute was not present. Somewhat counter to this was the finding that Blacks seemed to have performed relatively better than matched Whites when the response was Dependent on the Options as compared to when the responses were Independent of the Options. Findings for the other variables (i.e., Cannot be determined and Must/Could) were not significant.

Results for Order of Options were significant across ethnic groups: items were relatively more difficult for all three focal groups than for the matched White groups when the options were listed from Least (or lowest) to Greatest (or highest) as compared to when options showed no sequential ordering. These findings, while interesting, are not readily interpretable. Results pertaining to the absence or presence of Underlining in the stem were nonsignificant except perhaps for females, who seemed to have performed slightly worse when there was Underlining in the stem; the small number of items with this attribute and the small magnitude of the difference, however, call the significance of this finding into question.

Other Format Variables--TSWE: Item Format differences in TSWE are presented in Table 37. Earlier discussion mentioned that of the two item types in TSWE--Usage and Sentence Correction--focal groups performed relatively better on the latter, with its larger context and greater magnitude of choice (choose the "best" sentence rather than "identify where the error is"). When the two item types were pooled and

---

Insert Table 37 about here

---

the complexity of Sentence Structure was considered, females seemed to have performed somewhat better than matched males on the Complex as compared to the Simple sentences. (It is worth noting that here too relatively better female performance seemed to be associated with greater length, or larger context.) Differences in the ethnic groups were nonsignificant.

Results for the Error/No Error Key represented by the next variable were interesting in that females and Asian-Americans performed differentially better than matched reference groups on items in which the sentence presented was flawed (i.e., contained an Error that they were to identify) as compared to their relative performance (compared to the reference groups) on items that were supposed to be error-free; the reverse was true for Blacks. What this may mean is that females and Asian-Americans were more likely to see error or--perhaps more important--less likely to commit themselves to saying that a sentence was absolutely correct than were the matched reference groups. The findings indicate that Blacks, on the other hand, were more likely than matched Whites to be correct on items in which they committed themselves to saying that there was No Error (as in Usage) or that the sentence presented in the stem was the best (as in Sentence Correction).

Miscellaneous Format Variables: In the Item Format realm, results of a final set of variables are given in Table 38. Results are presented here across scales in

---

Insert Table 38 about here

---

an abbreviated format (i.e., without means and standard deviations and with levels of significance only where appropriate), since the findings were largely nonsignificant. As the table indicates, Key position in Mathematics had no effect, and Key position in Analogies and Sentence Completions had no effect. The position of the Key did result in marginally significant findings with males/females on Antonyms and significant findings for Blacks and Hispanics on Reading Comprehension. The pattern for Blacks and Hispanics on Reading Comprehension suggested relatively better performance for Whites on items with B keys and, to some degree, the reverse for items keyed as A, C, and E. Reasons for these findings are elusive. The use of the Roman Numeral Format had no effect in Reading Comprehension or in Mathematics; and the use of Negative terms in the stems of Reading Comprehension and Mathematics questions had negligible or no effect. Finally, in Reading Comprehension, there was no effect due to whether the stem was Closed (i.e., a complete sentence in and of itself) or Open (i.e., an incomplete question, completed in turn by each of the options). All of the foregoing variables were studied since they represent ways in which items differ from each other. As the results indicate, however, for the most part they had no discernible consistent effect on differential performance between groups.

Discussion and Summary

As is evident from this investigation, groups of students who achieve the same overall score on a test may not arrive at that score with the same pattern of responses. There are a multitude of factors that make some items relatively easier or harder for different groups of examinees, even after overall test score has been controlled. As was discussed earlier, some of these performance differences may reflect real gaps or deficiencies in the students' knowledge relative to the construct being measured by the test, while other differences may suggest that the item or items are measuring something extraneous to the construct being measured. The purpose of this investigation was to explore a myriad of potential factors in an effort to identify the variables that may warrant further experimental or policy-related review. In some instances, hypotheses in this investigation were drawn from past studies and our results either support or fail to support past findings. In other cases, the choice of variables to consider was pragmatic (based on existing test specifications) or, at the other extreme, speculative (based on the hunches of test developers or reviewers). Since this was an exploratory investigation, the goal was to be inclusive and perhaps to risk overidentification rather than to be definitive and perhaps to miss some findings that are suggestive of differences that may be masked by confounding variables or by too few items.

The factors that were considered in this investigation can be grouped into three main areas: (a) points tested, (B) item content, and (C) item format. Below, highlights of the investigation are summarized for each of the main reference/focal group comparisons. When considering these findings, it is important to keep in mind that differences do not reflect absolute group differences but rather relative performance discrepancies within item categories after the groups have been matched

for overall score. The highlights are followed by a discussion of the potential implications of these results and a consideration of the areas that warrant further research and review.

Highlights--Gender Differences

Verbal and TSWE

- On the Verbal scale, females performed relatively better than the matched group of males in Reading Comprehension and relatively less well on Antonyms and Analogies. This finding, suggesting that women seemed to do relatively better on item types with more context and relatively worse on those with little or no context, replicates earlier findings on the SAT. Similarly, on the TSWE Scale, females performed relatively better than matched males on Sentence Correction (which requires a choice of the best-written sentence) than on Usage (which requires a choice of if and where in a sentence an error occurs). Here, as in the Verbal scale, females seemed to do better when the task provided more context and less exact pinpointing.

- The subject matter content of the item represented a major factor in differential item performance for males and females. Results on all three Verbal discrete item types were significant and highly consistent. Females consistently performed relatively better than matched males on items in the fields of Aesthetics and Human Relationships and relatively less well than matched males on items in Science and in the World of Practical Affairs.

In Reading Comprehension, females, when compared to matched males, again performed relatively better in the field of Humanities and with Narrative passages

(a breakdown of Humanities) as compared to their relative performance on items with Science content; they also did relatively better in Social Science. In addition, females performed relatively less well than matched males on items based on Technical Science passages, particularly as compared to their relatively stronger performance (compared to matched males) on other Non-Science Reading Comprehension items.

In TSWE, females--as in the Verbal Scale--performed less well relative to matched males on sentences with Science content, particularly as compared to their relative performance on items dealing with Student Relevant concerns and Everyday Activities.

- Significant differences were found for another variable that may be related to or confounded by item subject matter content, that is, whether terminology in Analogy items is Abstract or Concrete. Drawing on the suggestion in the Wendler and Carlton study that women's performance might be enhanced on items in which the terms were Abstract rather than Concrete, the investigators examined Analogies on the Concrete-Abstract continuum. (Since the coding of the other Verbal item types would not have resulted in enough items in each cell, only Analogies were considered for this variable.) Results indicated that females tended to perform relatively better than matched males on Abstract Analogies, and relatively worse than matched males on those Analogies containing all Concrete terms. This finding leads to a research question for future study, that is, whether the relatively better performance of females in the areas of Humanities and Human Relations and their relatively weaker performance in the areas of Science and Practical Affairs are really completely a function of "discomfort" with or lack of exposure to the latter two fields or of whether these two fields tend to contain more Concrete rather than Abstract terms.

That is, to what extent is performance in subject matter areas associated with or confounded by the Abstract vs. Concrete nature of the terms used, and vice versa?

- Since Wendler and Carlton (1987) suggested that women might do somewhat worse with material that is upsetting or Controversial, this too was investigated. Taken together, the findings from these related variables lend tentative support to the Wendler and Carlton hypothesis that women may be more readily disrupted on some verbal item types when the content of the item is strongly Emotive or Controversial.

- Whether or not the item refers to People is another factor with consistent findings. With regard to Minority and Gender References and to references to People in general--studied in Sentence Completions and Reading Comprehension--females performed relatively better than matched males when People were referred to as compared to when there was no reference to People, even in those cases when ethnic group or Gender was not identifiable. Who is named or referred to did not seem to matter much; what did matter, in terms of relative male/female performance, was the presence or absence of People. TSWE showed a similar pattern: Females performed relatively better than matched males when People, especially females, were referred to in TSWE sentences.

Mathematics

- Males performed relatively better (compared to matched females) on Geometry and Geometry/Arithmetic items, while females performed relatively better than matched males on Miscellaneous and Arithmetic/Algebra items.

- Males performed relatively better when the item contained no variable (i.e., an unknown) while females performed relatively better when the item did contain a variable.

- Males found items with a Stimulus format (i.e., figure, graph, or table) relatively easier, while females performed relatively better than matched males when there was no Stimulus format.

- Males performed relatively better when the item called for a Computed solution, whereas females performed relatively better when the item called for a General solution. A somewhat contradictory finding was that females seemed to find Routine problems and those calling for mathematics Manipulation (lower-level cognitive processing) relatively easier than did matched males, whereas males seemed to find items requiring Higher-level cognitive processing relatively easier than did matched females.

- Females performed relatively better than matched males on items that were very much like the Curriculum rather than "real life" problems, whereas males tended to perform relatively better than matched females on the less routine "real life" problems. Also, because "real life" problems tended to be Word Problems, it is not surprising that males performed relatively better on other variables that could be associated with Word Problems (e.g., reading level, people references, length, etc.)

Highlights--White/Racial/Ethnic Background Comparisons

Verbal and TSWE

- Overall, considering the four Verbal item types, the only consistent pattern was that Whites performed relatively better than the focal groups on Analogies as compared to the other item types. Performance differences between item types for Whites and Asian-Americans were slight. Blacks performed less well relative to matched Whites on Analogies, and relatively better than matched Whites on Antonyms. Hispanics performed less well relative to matched Whites on Analogies and Sentence Completions and relatively better than matched Whites on Reading Comprehension.

- On the TSWE scale, Asian-Americans performed relatively better than matched Whites on Sentence Correction (which requires a choice of the best written sentence) than on Usage (which requires a choice of if and where in a sentence an error occurs). This trend was evidenced in all of the reference/focal group comparisons although it was significant only for the gender and White/Asian-American comparisons. Also on TSWE, with regard to elements of grammar and usage, results were significant for all reference/focal group comparisons. Of most interest are the findings that suggested that the area of Subject/Verb Agreement was consistently an area of relative difficulty for all focal groups compared to their respective matched groups and that detecting unwarranted Shifts in sentences and the related detecting of lack of Parallelism were consistently relatively less problematic for all focal groups.

- Subject matter content was found to represent an important factor in item performance. On the three discrete item types, six of the nine subject matter analyses resulted in significant findings. In each significant case, Science was associated with a performance discrepancy favoring the White reference groups whereas Human Relations was associated with relatively better performance by the focal groups. This relationship was found for Blacks on all three discrete item types, for Hispanics on Antonyms and Analogies, and for Asian-Americans on Analogies.

In Reading Comprehension, the trends apparent in the discrete Verbal item types did not hold consistently. No significant results were obtained for Blacks or Hispanics. Unlike their performance on Analogies, in Reading Comprehension, Asian-Americans tended to perform relatively better on Science passages, particularly Technical ones, than on Humanities passages. This disjunction in the Asian-American group is hard to explain. Possibly, language problems were either exacerbated in the more literary or general passages or modified by technical language, which might be more familiar and more accessible to students from an Asian-language background.

In TSWE, some of the same results were seen for Asian-Americans and the other ethnic groups as in the Verbal discretes. Asian-Americans and Hispanics performed less well when compared to matched reference groups on sentences with Science content; Asian-Americans and Hispanics performed relatively better than matched Whites on Social Science; and Hispanics tended to perform relatively better on items dealing with Everyday Activities and Student Relevant concerns. Results for Blacks on TSWE subject matter breakdowns, while not significant, were nonetheless interesting in that they were consistent with the tendency for focal groups to experience relatively more difficulty than matched Whites with items with a Science context.

• When Analogies were examined on the Concrete-Abstract continuum, as had been hypothesized there was a clear trend for both females and ethnic groups to perform relatively better than matched referenced groups on Abstract Analogies, and relatively less well on those containing all Concrete terms. Relative performance on Analogies containing a Mix of Abstract and Concrete terms fell somewhere in between. Results for ethnic group comparisons suggested larger relative discrepancies than those for gender. As with females, one question for further study is whether the relatively better performance of Blacks, and Hispanics in the areas of Humanities and Human Relations and their relatively poorer performance in the areas of Science and Practical Affairs is a function of "discomfort" with, or lack of exposure to, the latter two fields or whether items in these two fields tend to contain more Concrete than Abstract terms.

• Another item content variable dealt with the absence or presence of Graeco/Latin word origins. For Analogies, the results for Asian-Americans, Blacks, and Hispanics were significant and indicated better performance relative to matched Whites on Latinate items, less of a performance discrepancy on mixed items, and relatively poorer performance (compared to matched Whites) on items using words without Classical origins. The Latinate variable yielded no significant results for Antonyms.

• When Analogies were examined for the absence or presence of Homographs (words spelled the same but having different meanings, as in "bear," "bark," "press"), significant findings resulted for all three reference/ethnic group comparisons. The presence of two Homographs was associated with a larger reference/ethnic group performance discrepancy (favoring Whites) than was the

absence of Homographs for all groups. One hypothesis is that ethnic group performance was more disrupted than matched White group performance by the potentially confusing appearance of words that looked like but were different from more commonly appearing words. Results on this variable for Antonyms were nonsignificant.

- In Reading Comprehension the Reference to or naming of Minority members tended to be associated with relatively better performance by focal group members. Minority groups also tended to perform better relative to matched Whites when Gender References were made or when Females or Males were named than when there were no People or no one named. In addition, on Analogies, there was a consistent pattern for Asian-Americans, Hispanics, and Blacks to perform relatively better than matched Whites when People were referred to than when they were not.

- Another format variable, whether options are presented Vertically or whether they are presented Horizontally, in "Wraparound" fashion, yielded striking results for Analogies. All ethnic groups performed less well than matched Whites when items appeared in a Wraparound fashion as compared to their relative performance on items when options were presented in a Vertical format; this finding has been replicated using similar methodology for a White/Black comparison on the Graduate Record Examinations.

- A further variable that was studied dealt with how well examinees could adjust to a new item type or switch from one item type to another. When First Appearance of an item type (or its subsequent First Reappearance) was compared to "other items," that is, items that follow the first one in a test section, First

Appearances were associated with unexpectedly poorer performance by ethnic groups relative to matched groups of Whites. One hypothesis to be explored is whether there was a "jolt by the unfamiliar" effect, which disrupted ethnic group performance more than reference group performance. An alternative hypothesis is that there is some other confounding characteristic associated with the initial items in each section.

- Also considered was whether or not the terms in the stem (or question part) of an Analogy come from the same domain as the options (in which case there is a Vertical relationship and the item is said to be Overlapping). When the key was related Vertically, as compared to when there was no Vertical relationship, ethnic groups performed less well than matched Whites. Further, when <u>any</u> distractor had a Vertical relationship to the stem, as compared to items without a Vertical relationship, the performance of Blacks and Hispanics was relatively weaker than that of matched Whites. The relatively poorer performance of focal groups (significant for all of the minorities) on the Overlapping category of the Independent/Overlapping variable simply confirms the foregoing, since Overlapping indicates that the stem and the key come from the same domain. With regard to the results for all three of the variables here, one hypothesis is that the confusion or distraction caused by options that are closely related to the stem in subject matter may have disrupted focal group performance more so than reference group performance.

- Although several item format differences in TSWE were evaluated, one that is particularly interesting, although difficult to interpret, is the Error/No Error Key. Asian-Americans performed differentially better than matched Whites on items in which the sentence presented was flawed than on items that were supposed to be

error-free; Blacks significantly did the reverse. What this may mean is that Asian-Americans (on the average) are more likely to see error or--perhaps more important--less likely to commit themselves to saying that a sentence is absolutely correct than are the matched Whites. The findings indicate that Blacks, on the other hand, were more likely to be correct on items in which they committed themselves to saying that there was No Error (as in Usage) or that the sentence presented in the stem was the best (as in Sentence Correction).

Mathematics

- Asian-Americans and Hispanics performed relatively better than matched reference groups on Geometry items--Geometry/Algebra and Geometry/Arithmetic, while Whites performed relatively better than these focal groups on Arithmetic and mathematics items categorized as Miscellaneous.

- Like females, all the ethnic groups performed relatively better than matched Whites on items that contained a variable, whereas Whites performed relatively better on items on which no variable was present.

- Asian-Americans and Hispanics seemed to do relatively better than matched groups of Whites on items with a spatial component. These groups did relatively better on items involving geometry and estimation and also relatively better on items containing Figures, as compared to their relative performance on the other item categories within these variables. Items involving Figures (particularly when the figure was not provided) consistently were associated with a performance discrepancy favoring Asian-Americans and Hispanics, as compared to items not involving Figures.

- Similar to the findings for females was the finding that all ethnic groups performed relatively better than matched Whites on mathematics items that were very much like the Curriculum and not "real world" problems, whereas matched White comparison groups performed relatively better on the "real world" problems. These "real world" problems tended to be Word Problems. Related differences were found for the other variables that would be associated with Word Problems (i.e., reading level, length, and references to people).

- Asian-Americans tended to perform relatively better than matched Whites on mathematics problems involving mathematics Manipulation and Routine problem solving as compared to their poorer performance relative to matched Whites on mathematics problems involving Higher-level thinking skills.

- Blacks, relative to matched Whites, performed relatively better when Arithmetic/Algebra was required as compared to their relative performance when Arithmetic alone was required. Unlike the Asian-Americans and Hispanics, but similar to females, Blacks performed relatively better than matched Whites on items without a Stimulus format and those not involving Figures as compared to their relative performance when Figures were involved.

Implications

At this point, one could ask, "Now what?" There are many interesting findings, some of which support past related research or seem to be associated with patterns of performance, as, for example, the hypothesized "jolt effect"

by the new or by controversial material. What is the value of identified differences between groups matched on overall performance?

Since tests are an important part of educational decisions, we need to understand how tests work and what they really measure. One question that could be raised is one of construct validity. If different groups of examinees with the same overall score arrive at the score in very different ways, one could question whether the same construct is being evaluated for both groups. Evidence from this investigation does not support the notion that different constructs are being measured, but rather that there may be stimuli associated with how an item is presented (content or format) that (on the average) differentially affect the performance of focal or reference group members.

Obviously, more research is needed to confirm the findings and to disentangle any results that may be confounded (as, for example, the Concrete-Abstract continuum vs. item subject matter context). If, however, the findings discussed here are confirmed, what are the implications? Implications exist for a variety of educators. First, there are implications for curriculum developers, since significant differences in points tested (which occur far more frequently in the Mathematics and TSWE scales than in the Verbal scale) may well point to areas that are underemphasized in the curriculum or to areas in which increased experience or remediation or even different methods of teaching might lead to improvement in the performance of focal groups. Second, there are major implications for test developers and test sponsors, who need to consider and reconsider these variables in deciding both what should be in tests and in what quantity. That is, given, for example, the seemingly differential effect of discrete Science items on all

focal groups, decisions need to be made regarding what kind of Science items go into tests, in what quantity they go into tests, and also--in order to be equitable to test-takers at every test administration--how these specifications can be monitored and controlled to ensure an evenness of both positive-impact and negative-impact characteristics from test form to test form.

Further, implications exist for test assemblers and test sponsors in configuring tests and in deciding how tests are organized, how sample items and directions are presented, and, perhaps, even in deciding how best to develop and use materials for test preparation. Given the possible "jolt" effect of new test material reported on here, familiarizing all groups with test material before the test may take on even greater importance than formerly.

In addition, continued consideration needs to be given--by educators at all levels and especially by test developers--to the several issues surrounding the influence of language on the performance of Asian-Americans and Hispanics. Although this study investigated only the results of examinees for whom English was self-reported to be the best language, there is nevertheless evidence (e.g., Word Problems in Mathematics, Homographs in Verbal) to suggest that less-than-perfect familiarity with the English language may be differentially influencing the performance of even other examinees. To what extent should relative lack of fluency--perhaps temporary--be allowed to influence, or be removed from influencing, test scores?

Research Directions

One purpose of this investigation was to explore a broad spectrum of item characteristics in order to identify those that seemed to be associated with differential item functioning and that warranted further experimental review. Several avenues of research seem called for. First, many categories emerged that overlapped such that the results were confounded and it was not possible within the limitations of this investigation to disentangle the findings. For example, with Analogies, the previously discussed Concrete/Abstract variable seemed to be confounded by the item Subject Matter variable. That is, the items in the Science Subject Matter area tended to be Concrete. In Science items that exhibit DIF, then, which of these overlapping elements (or perhaps some combination of elements) is responsible? A better understanding of the impact of each of the factors in such overlapping categories would be useful for educators and test developers. Such an understanding could come from developing and pretesting items that separated these characteristics (e.g., Science items that are Abstract) or by further investigating past tests to locate a sufficient number of items in which the variables are not confounded.

Future studies should also consider the relationship between item categories and item difficulty. For the most part, items in the various categories identified in this study were distributed across different difficulty levels. For example, in the examination of data for each of the combined matched groups, it was found that the difficulty level (or average delta) of Science items was comparable to the difficulty level of Aesthetics/ Philosophy items. This was not the case, however, for First Appearance items.

Often the first item in a section is a relatively easy item; the possible "jolt effect" noticed with First Appearance of an item, therefore, may in part be explained by the general practice of placing easier items earlier in a test section. Studies unconfounding format or subject matter, on the one hand, and difficulty level, on the other hand, would contribute to knowledge about the possible contribution of item difficulty per se to differential performance by groups.

Further, methodical manipulation of experimental tests should be routinely undertaken, in which the various hypotheses that grow out of this and other studies are systematically tested in order to determine which of the hypotheses do indeed hold up. In these tests, items would be built that specifically and in large numbers contain the categories thought to contribute to differential performance. With these categories both controlled and represented in sufficient numbers, one could better separate those elements that contribute to differential performance from those that are merely artifactual.

Finally, an interesting and potentially useful series of studies would involve evaluating the impact of manipulating the test--perhaps even while maintaining current test specifications--in an attempt to slant the test in favor of a particular focal group. Using the results of this and similar investigations, one could devise one or more experimental pretests that took into account the factors that seem to be associated with differential reference and focal group performance. While it is true that one could probably develop such a test (e.g., a Verbal test that favored females by concentrating on Reading Comprehension Human Relations and Aesthetics/Philosophy items), it is less certain that the test would

measure the domains thought to be relevant for success in college. By starting with the test specifications currently in use, however, one could test the limits of the current system. Should the limits prove to be too constraining, it would be useful to determine to what extent specifications would need to change in order to significantly decrease differential performance between groups.

REFERENCES

Alderman, D. L. (1982). Language proficiency as a moderator variable in testing academic aptitude. Journal of Educational Psychology, 74, 580-587.

Bleistein, C. A., & Wright, D. (1987). Assessment of unexpected differential item difficulty for Asian-American examinees on the Scholastic Aptitude Test. In A. P. Schmitt & N. J. Dorans (Eds.). Differential item functioning on the Scholastic Aptitude Test. (ETS Research Memorandum No. RM-87-1). Princeton, NJ: Educational Testing Service.

Bleistein, C. A., & Wright, D. (1985). Assessing unexpected differential item performance of Oriental candidates and of White candidates for whom English is not the best language on SAT Form 3FSA08 and TSWE Form E47. (ETS Statistical Report No. SR-85-123). Princeton, NJ: Educational Testing Service.

Carlton, S. T., & Marco, G. L. (1982). Methods used by test publishers to "debias" standardized tests: Educational Testing Service. In R. A. Berk (Ed.). Handbook of methods for detecting test bias. Baltimore, MD: Johns Hopkins Press.

Clark, M. J., & Grandy, J. (1984). Sex differences in the academic performance of Scholastic Aptitude Test takers. (College Board Report No. 84-8). (ETS Research Report No. 84-43). Princeton, NJ: Educational Testing Service.

Donlon, T. F. (1981). The SAT in a diverse society: Fairness and sensitivity. The College Board Review, No. 122. 16-21,30-32.

Donlon, T. F., Ekstrom, R. B., Lockheed, M. E., & Harris, A. M. (1977). Performance consequences of sex bias in the content of major achievement batteries. (Project Report 77-11). Princeton, NJ: Educational Testing Service.

Doolittle, A. E., & Cleary, T. A. (1987). Gender-based differential item performance in mathematics achievement items. Journal of Educational Measurement, 24, 157-166.

Dorans, N.J. (1982). Technical review of SAT item fairness studies: 1975-1979. (ETS Statistical Report No. SR-82-90). Princeton, NJ: Educational Testing Service.

Dorans, N.J., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. Journal of Educational Measurement, 23, 355-368.

Freedle, R., Kostin, I., & Schwartz, L. (1987) . A comparison of strategies used by Black and White students in solving SAT verbal analogies using a thinking aloud method and a matched percentage-correct design. (ETS Research Report No. RR-87-48). Princeton, NJ: Educational Testing Service.

Holland, P. W., & Thayer, D. T. (1986). Differential item performance and the Mantel-Haenszel procedure. (ETS Program Statistics Research Technical Report No. 86-69). (ETS Research Report No. 86-31). Princeton, NJ: Educational Testing Service.

Hunter, R. V., & Slaughter, C. D. (1980). Educational Testing Service sensitivity review process. Princeton, NJ: Educational Testing Service.

Leary, L. F., & Dorans, N. J. (1982). The effects of item rearrangement on test performance: A review of the literature. (ETS Research Report No. RR-82-30). Princeton, NJ: Educational Testing Service.

Linn, M. C., De Bendictis, T., Delucchi, K., Harris, A. M., & Stage, E. (1987). Gender differences in National Assessment of Educational Progress Science items: What does "I don't know" really mean? Journal of Research in Science Teaching, 24, 267-278.

Linn, R. L. (1973). Fair test use in selection. Review of Educational Research, 43, 139-161.

Linn, R. L., & Harnisch, D. L. (1981). Interaction between item content and group membership on achievement test items. Journal of Educational Measurement, 18, 108-118.

O'Neill, K., McPeek, W. M., & Wild, C. L. (in progress). Differential item functioning on the Graduate Management Admission Test.

O'Neill, K., Wild, C. L., & McPeek, W M. (in progress). Differential item functioning on the Graduate Record Examinations General Test.

Pearlman, M. A. (1987). Trends in women's total score and item performance on verbal measures. Paper presented at the Annual Meeting of the American Educational Research Association.

Pennock-Roman, M. (1988). The status of research on the Scholastic Aptitude Test (SAT) and Hispanic students in postsecondary education. (ETS Research Report No. RR-88-36). Princeton, NJ: Educational Testing Service.

Rogers, H. J., & Kulick, E. (1987). An investigation of unexpected differences in item performance between Blacks and Whites taking the SAT. In A.P. Schmitt & N.J. Dorans (Eds.). Differential item functioning on the Scholastic Aptitude Test. (ETS Research Memorandum No. RM-87-1). Princeton, NJ: Ed. Testing Service.

Rosser, P. (1987). Sex bias in college admissions tests: Why women lose out (2nd ed.). Cambridge, MA: National Center for Fair and Open Testing.

Schmitt, A. P. (1988). Language and cultural characteristics that explain differential item functioning for Hispanic examinees on the Scholastic Aptitude Test. Journal of Educational Measurement, 25, 1-13.

Schmitt, A.P., & Dorans, N.J. (1988). Differential item functioning for minority examinees on the SAT. (ETS Research Report No. RR-88-32). Princeton, NJ: Educational Testing Service.

Welch, C. J., & Doolittle, A. E. (1988). Gender-based differential item performance in English usage items. Paper presented at the Annual Meeting of the American Educational Research Association.

Wendler, C. L. W., & Carlton, S. T. (1987). An examination of SAT verbal items for differential performance by women and men: An exploratory study. Paper presented at the Annual Meeting of the American Educational Research Association.

Wheeler, P., & Harris, A. M. (1981). Comparison of male and female performance on the ATP Physics Test. NY: College Entrance Examination Board.

Wild, C. L., & Durso, R. (1979). Effect of increased test-taking time on test scores by ethnic group, age, and sex. (Graduate Record Examinations Board Research Report No. GREB-76-6R). Princeton, NJ: Educational Testing Service.

APPENDIX A

CODING CATEGORIES: VERBAL
SAT VARIABLES COMMON TO ALL VERBAL ITEM TYPES

MINORITY STIMULUS Column 15

1 = Black Americans	Stimulus refers to Black Americans
2 = Hispanic Americans	Stimulus refers to Hispanic Americans
3 = Native Americans	Stimulus refers to Native Americans
4 = Asian Americans	Stimulus refers to Asian Americans
5 = Third-World Black	Stimulus refers to Third-World Blacks
6 = Third-World Hispanic	Stimulus refers to Third-World Hispanics
7 = Third-World Asian	Stimulus refers to Third-World Asians
8 = Nonminority Ethnic	Stimulus refers to other ethnic groups
9 = General	Stimulus refers to people of no specified ethnic origin
0 = Nothing	Stimulus does not refer to people

GENDER REFERENCE IN STIMULUS Column 16

1 = Female	Stimulus refers to females only
2 = Male	Stimulus refers to males only
3 = Mixed	People referred to in stimulus are unidentified as to whether they are male or female (such as teachers, they, we, you, students)
5 = Neutral	Stimulus does not refer to people

NEGATIVE ITEM (No/Except) Column 17

1 = Negative Stem	Use of "NOT", "CANNOT", "EXCEPT", "LEAST", "INCORRECT", "FALSE" etc., in stem
2 = Positive stem	Do not use "NOT", etc., in stem

ROMAN NUMERAL FORMAT Column 18

1 = Roman	Involves Roman numeral format
2 = No-Roman	Does not involve Roman numeral format

ITEM TYPE Column 19

1 = Sentence Completion
2 = Analogies
3 = Reading Comprehension
4 = Antonyms

EMOTIVE QUALITY Column 20

0 = Sentence refers to neutral or pleasant subject matter.

1 = Sentence refers to strongly upsetting subject matter (e.g., evil, fire, flood, nuclear war.) NOTE: This list will be expanded during coding. This does not include argumentative or inflammatory subject coding; rather it refers to questions which have a negative impact or an overall tone of a depressing nature. The word strongly is the clue here.

2 = Can't Decide (NOTE: Use can't decided as a flag or signal that the coding descriptions need clarifying or that another opinion is needed. Ultimately, all items should fit into the coding categories.)

ITEM FORMAT Refers to the form used to set up the item or the way the item appears in the test.

Analogies	Column 26
Antonyms	Column 30
Sentence Completions	Column 40
Reading Comprehension	Column 47

1 = Type 1
(each choice on separate lone or arranged vertically)

CHOIR: SINGER::
(A) election: voter
(B) anthology: poet
(C) cast: actor
(D) orchestra: composer
(E) convention: speaker

2 = Type 2

WATER: SWIM:: (A) grass: grow (B) knot: tie (C) plan: implement (D) flood: damage (E) snow: ski

(choices run together or arrange horizontally on two lines or more)

SAT VARIABLES STUDIED IN VERBAL: ANALOGIES

(See also "Variables Common to All Verbal Item Types.")

SUBJECT CONTENT Column 21

1 = Aesthetic - Philosophy	Includes art, architecture, drama, literature, music, religion, philosophy
2 = Practical Affairs	Includes sports, economics, business, communications, politics, transportation, other special sciences
3 = Science	Includes mathematics, medicine, technology, applied science, agriculture, manual arts
4 = Human Relationships	Includes emotions, character analyses, interpersonal relationships, general psychology
5 = Mixed/Overlapping	Includes a mixture of content or content that overlaps 2 or more categories

KINDS OF ANALOGIES Column 22

1 = Concrete	An analogy is usually classified as concrete if the terms in stem and key refer to entities that can be perceived by one or more of the primary senses (sight, hearing, smell, touch, and taste).
2 = Mixed	An analogy is classified as mixed if some, but not all, of the items in stem and key refer to entities that can be perceived by one or more of the primary senses.
3 = Abstract	An analogy is classified as abstract if none of the four terms in stem and key refers to entities that can be perceived by one or more of the primary senses.

INDEPENDENT/OVERLAPPING Column 23

1 = Independent

An analogy is independent if neither of the terms in the stem has either of the relationships listed below (class/subclass, subclass/class) with a term in the key.

Example:

CUMULUS: CLOUD::
(A) lake: ocean
(B) carnivore: meat
(C) glacier: blizzard
(key) (D) evergreen: pine
(E) evening: daylight

2 = Overlapping

An analogy is overlapping if one or both of the terms in the stem has/have either a class/subclass or a subclass/class relationship to one or both of the terms in the key.

Example:

REFINE: PETROLEUM::
(A) consume: fuel
(key) (B) smelt: ore
(C) prospect: uranium
(D) blend: alloy
(C) import: rubber

TAXONOMY OF SEMANTIC RELATIONSHIP Columns 24-25

This refers to the nature of the relationship between the words in the stem of an analogy item. This relationship can define the type of association that needs to be made in order to correctly identify the option with the same relationship (key).

The Chaffin/Peirce taxonomy consists of several families of relationships that were described by Chaffin and Peirce (1986) as follows:

0 = Can't decide

1 = Class Inclusion

One word names a class that includes the entity named by the other word. (e.g., faculty: teachers)

2 = Part - Whole: Positive

One word names a part of the entity named by the other word. (e.g., car: engine)

3 = Part - Whole: Negative

One word names a part of the entity that can never be part of the entity named by the other word. (e.g., tundra: tree; perfection: fault)

4 = Similar - Degree Relationships	One word represents a different degree of the object, action, or quality represented by the other word. (e.g., enthusiasm: fervor; eating: gluttony)
5 = Similar - Other Relationships	One word represents a different form of the object, action, or quality represented by the other word. (e.g., listen: eavesdrop; rake: fork)
6 = Attribute	One word names a characteristic quality, property, or action of the entity named by the other word.
7 = Contrast	One word names an opposite or incompatible of the entity named by the other word.
8 = Nonattribute	One word names a quality, property, or action that is characteristically <u>not</u> an attribute of the entity named by the other word. (That which is first lacks the second quality.)
9 = Case Relation	One word names an action which the entity named by the other word is usually involved in, or both words name entities that are normally involved in the same action in different ways, e.g., as agent, object, recipient, or instrument of the action.
10 = Cause/Purpose	One word represents the cause, purpose, or goal of the entity named by the other word, or the purpose or goal of using the entity named by the other word. (You do first to the second. You do first to get rid of the second, etc.)
11 = Space / Time	One word names a thing or action that is associated with a particular location or time named by the other word.
12 = Representation	One word names something that is an expression or representation of, or a plan or design for, or provides information about, the entity named by the other word.
13 = Other	

ITEM FORMAT Column 26

Refers to the form used to set up the item or the way the item appears in the test.

1 = Type 1
(each choice on separate lone or arranged vertically)

CHOIR: SINGER::
(A) election: voter
(B) anthology: poet
(C) cast: actor
(D) orchestra: composer
(E) convention: speaker

2 = Type 2

WATER: SWIM:: (A) grass: grow
(B) knot: tie (C) plan: implement
(D) flood: damage (E) snow: ski

(choices run together or arrange horizontally on two lines or more)

PARTS OF SPEECH IN ANALOGY STEMS Column 27
(Keys have the same parts of speech as the stems)

1 = noun: noun
2 = noun: adjective or adjective: noun
3 = noun: verb or verb: noun
4 = verb: verb
5 = verb: adjective or adjective: verb
6 = adjective: adjective

KIND OF LANGUAGE (Technical / Non-Technical) Column 28

This refers to the use of technical language (i.e., part of the jargon of a field) as opposed to general, everyday, accessible language.

0 - Stem, key, distractors contain no technical terms
1 = Stem only, contains 1 or 2 technical terms
2 = Key only, contains 1 or 2 technical terms
3 = One or more distractors only, contains 1 or 2 technical terms
4 = Stem and key only, contain 1 or 2 technical terms.
5 = Stem and/or key and one or more distractors containing 1 or 2 technical terms.
6 = Can't decide.

LATINATE LANGUAGE Column 29

0 = No term in stem or key is Latinate/Greek.
1 = Stem and key have all Latinate/Greek terms.
2 = Stem and key have mixed Latinate/Greek and other (e.g., Anglo-Saxon) terms.

SCIENCE Column 30

Refers to the main or predominating category of science found in the item (stem and options).

0 =	No Science Content	
1 = Biology	Biology of animals, human anatomy	
2 = Botany	Biology of plants	
3 = Physical	Includes physics, earth science, chemistry, astronomy	
4 = Applied Science and Technology	Includes agriculture, transportation, computer science, health, or medicine	
5 = Mathematics		
6 = Mixed/Overlapping	Includes a mixture of science content or science content that overlaps two or more science categories	

IDEA ASSOCIATION

This refers to a non-analogical answering strategy where the relationship between the terms in the stem may NOT be considered when selecting the correct option. Instead, the examinee appears to use idea association (as in Physician: Hospital:: Nurse: Patient). Idea association analogies are ones in which some words in the options belong to the same general area of discourse as one or both words in the stem. An up and down or vertical* strategy is used rather than the horizontal or across (XXX:XXX:: (A) XXX:XXX) strategy used to correctly solve an analogy item. In idea association strategy, each word in the stem is looked at individually and associated with words in the option rather than looking at them as a pair with a distinct relationship. Thus, in the example below, instead of choosing the key (D) with the correct relationship, (B) might be chosen because of the idea association between pottery and wheel.

Example: SHARD: POTTERY
(A) flint: stone
(B) flange: wheel
(C) cinder: coal
(D) fragment: bone
(E) tare: grain

* *Alicia P. Schmitt and Carole A. Bleistein, Factors Affecting Differential Item Functioning for Black Examinees on SAT Analogy Items (1986).*

IDEA ASSOCIATION BETWEEN STEM AND KEY Column 31

Enter the presence or absence of all obvious idea association between the words in the stem and words in the key.

0 = No vertical relationship between stem and key

1 = Vertical relationship between stem and key

IDEA ASSOCIATION BETWEEN STEM AND DISTRACTORS Column 32

Enter the presence or absence of an obvious idea association between the words in the stem and the words in one or more distractor(s).

0 = No vertical relationship between stem and distractor(s)

1 = Vertical relationship between stem and distractor(s)

HOMOGRAPHS

Homographs are words that are spelled the same, but have significantly different meanings or pronunciations (as defined in Webster's Ninth New Collegiate Dictionary) which are accessible, common, ordinary, or plausible. (For example: bark, table, bad, temper, clip.) All words should be checked in the dictionary to be sure that significant homographs are not overlooked.

HOMOGRAPHS IN STEM Column 33

0 = Stem does not contain significant homograph
1 = Stem contains 1 significant homograph
2 = Stem contains 2 significant homographs

HOMOGRAPHS IN KEY Column 34

0 = Stem does not contain significant homograph
1 = Stem contains 1 significant homograph
2 = Stem contains 2 significant homographs

HOMOGRAPHS IN DISTRACTORS Column 35

0 = No distractor contains a significant homograph
1 = One or more distractors contain 1 or more significant homographs

KEY - ANALOGIES Column 36

1 = A
2 = B
3 = C
4 = D
5 = E

CONTROVERSIAL SUBJECT MATTER - Column 37

Refers to subjects of a controversial or inflammatory nature -- argumentative or current topics in the news. Anything which can be argued or debated. Often are sociopolitcal in nature.

1 = Yes	Attribute is present
2 = No	Attribute is not present

SAT VARIABLES STUDIED IN VERBAL: ANTONYMS

(See also "Variables Common to All Verbal Item Types.")

SUBJECT CONTENT Column 21

1 = Aesthetic/Philosophy	Includes art, architecture, drama, music, religion, literature, philosophy
2 = Practical Affairs	Includes sports, communications, politics, transportation, government, business, economics
3 = Science	Includes technology, applied science, agriculture, manual arts
4 = Human Relationships	Includes emotions, character analyses, interpersonal relationships
5 = Mixed/Overlapping	Includes a mixture of content or content that overlaps two or more categories

KINDS OF ANTONYMS Column 22

1 = General Definition Antonym	A general definition antonym is one in which none of the distractors is related in meaning to the intended key.
2 = Fine Distinction Antonym	A fine distinction antonym is one in which at least one of the distractors is related in meaning to the intended key, but is not as complete or exact as the intended key. (Closely related distractor that is incomplete opposite of the stem)

PART OF SPEECH Column 23

1 = Verb The stem and options are verbs.
2 = Noun The stem and options are nouns.
3 - Adjectives or adverbs The stem and options are adjectives or adverbs.

SINGLE WORD/PHRASES Column 24

1 = All of the options are single words.
2 = All of the options are short phrases.
3 = Some of the options are single words and some of them are phrases.

HOMOGRAPHS Column 25

Homographs are words that are spelled the same, but have significantly different meanings or pronunciations (as defined in Webster's Ninth New Collegiate Dictionary) which are accessible, common, ordinary, or plausible. (For example, bark, table, bad, temper, clip.) Stem and key are to be checked in the dictionary.

0 = No obvious homographs in stem, key or distractors.
1 = Obvious homograph in stem only.
2 = Obvious homograph in key only.
3 = Obvious homograph(s) in distractor(s) only.
4 = Obvious homographs in stem or key.
5 = Obvious homographs in stem and/or key and distractors(s).

KIND OF LANGUAGE (Technical/Non-Technical) Column 26

0 = Stem, key distractors contain non-technical terms.
1 = Stem only, contains technical term.
2 = Key only, contains technical term.
3 = One or more distractors only, contain technical term.
4 = Stem and key only, contain technical terms.
5 = Stem and/or key, and one or more distractors, contain technical term(s).
6 = Can't decide.
(NOTE: Use only as a flag or signal that the coding descriptions need clarifying or that another opinion is needed. Ultimately, all items should fit into the coding categories.)

LATINATE LANGUAGE Column 27

0 = No Latinate/Greek language in stem or key.
1 = Stem and key have Latinate/Greek terms.
2 = Stem and key have mixed Latinate/Greek and other terms.

ABSTRACT VERSUS CONCRETE SPECIFIC LANGUAGE Column 28

1 = Stem and key are concrete (i.e., contain words perceivable by one or more of the five senses).
2 = Stem and key are mixed.
3 = Stem and key are abstract.
4 = Not applicable.

KEY - ANTONYMS (Refers to the option which is the intended key) Column 29

1 = A
2 = B
3 = C
4 = D
5 = E

ITEM FORMAT Column 30 (See p. 68)

SAT VARIABLES STUDIED IN VERBAL: SENTENCE COMPLETIONS

(See also "Variables Common to All Verbal Item Types".)

LENGTH OF STEM (number of words) Columns 21-23

The actual word count was entered on the coding sheet for each item. (Hyphenated words, numbers, and Roman numerals were counted as one word.)

LENGTH OF OPTIONS (A through E) Columns 24-26

1 = Each option is a single word.
2 = Each option is a pair of single words.
3 = Each option is a phrase, either single or in pairs.
4 = Options are single words paired with a phrase.

NUMBER OF BLANKS (refers to the number of blanks in the stem) Column 27

1 = One blank
2 = Two blanks

KIND OF LANGUAGE (Technical/Non-Technical) Column 28

This refers to the use of technical language as opposed to general, everyday, accessible language.

0 = Stem, key, distractors contain no technical terms.
1 = Stem only, contains 1 or two technical terms.
2 = Key only, contains 1 or two technical terms.
3 = One or more distractors only, contain 1 or 2 technical terms.
4 = Stem and key only, contain 1 or 2 technical terms.
5 = Stem and/or key and one or more distractors, contain 1 or 2 technical terms.
6 = Can't decide.

LATINATE LANGUAGE Column 29

0 = No terms in stem or key are Latinate/Greek.
1 = Stem and key have all Latinate/Greek terms.
2 = Stem and key have mixed Latinate/Greek and other (e.g., Anglo-Saxon) terms.
3 = Can't decide.
4 = Not applicable

CONTROVERSIAL SUBJECT MATTER Column 30

Refers to subjects of a controversial or inflammatory nature -- could be argumentative or in the current news. Anything which can be argued or debated. Includes topics which are sociopolitical in nature.

1 = Yes	The attribute is present.
2 = No	The attribute is not present.

SOCIALLY RELEVANT Column 31

A socially relevant sentence is a sentence whose content concerns an aspect or issue of contemporary society that is related to social justice or to political, legal, or economic equality.

1 = Yes	The item is socially relevant.
2 = No	The item is not socially relevant.

SUBJECT CONTENT Column 32

1 = Aesthetic/Philosophy	Includes art, architecture, drama, music, religion, literature, philosophy
2 = Practical Affairs	Includes sports, communications, politics, transportation, government, business, economics
3 = Science	Includes technology, applied science, agriculture, manual arts
4 = Human Relationships	Includes emotions, character analyses, interpersonal relationships
5 = Mixed/Overlapping	Includes a mixture of content or content that overlaps two or more categories

NAMED I - MINORITY IN ITEM Column 33

Refers to a person or persons specifically named or referred to)

0 = No one named or referred to
1 = Black American named or referred to
2 = Hispanic American named or referred to
3 = Native American named or referred to
4 = Asian American named or referred to
5 = Third-world Black named or referred to
6 = Third-world Hispanic named or referred to
7 - Third-world Asian named or referred to
8 = Nonminority ethnic group named or referred to
9 = General or unidentifiable

NAMED II - GENDER IN ITEM Column 34
(Refers to a person or persons specifically named or referred to.)

0 = No one named or referred to
1 = Female named or referred to
2 = Male named or referred to
3 = Mixed named or referred to
4 = General or unidentifiable

LOCATION OF BLANK(S) IN SENTENCE Column 35

1 = First (or only) blank appears before subject of sentence, or first blank is the subject.
2 = First (or only) blank appears after subject but before main verb (but before end of first clause if more than one clause). Or, first blank is main verb.
3 = First (or only) blank appears after first clause.
4 = First blank appears after both subject and main verb, in first clause.

PARTS OF SPEECH IN OPTIONS* Column 36

1 = One noun
2 = One adjective
3 = One verb
4 = Two nouns (in 2-blank sentence)
5 = Two adjectives (in 2-blank sentence)
6 = Two verbs (in 2-blank sentence)
7 = Other parts of speech
8 = Mixed parts of speech (in 2-blank sentence)

MEANS OF FINDING ANSWER Column 37

(Refers to what appears to be the most appropriate or likely strategy.)

1 = Answer comes to mind after question is read and student then searches for closest counterpart among options.
2 = Student must look at all options before formulating or identifying answer.

SENTENCE STRUCTURE Column 38

1 = Sentence is simple.
2 = Sentence is compound.
3 = Sentence is complex, with 1 dependent clause.
4 = Sentence is complex, with 2 or more dependent clauses.
5 = Sentence is compound-complex.

KEY - SENTENCE COMPLETIONS Column 39

1 = A
2 = B
3 = C
4 = D
5 = E

ITEM FORMAT Column 40 (See p. 68)

** Excluding modifying articles and prepositions*

SAT VARIABLES STUDIED IN VERBAL: READING COMPREHENSION
(See also "Variables Common to All Verbal Item Types".)

LENGTH OF STIMULUS Column 21-23

Actual word* counts were entered.

LENGTH OF STEM Column 24-26

Actual word* counts were entered.

LENGTH OF OPTIONS Columns 27-29

Actual word* counts were entered.

OPTION FORMAT Column 30

1 = Very short phrase (1-4 words)
2 = Longer phrase
3 = Sentence

QUESTION FORMAT Column 31

1 = Open — An open question is one in which the actual question being asked is not a complete sentence and is completed by the options.

2 = Closed — A closed question is one in which the actual question being asked is a complete sentence and the options are independent units (words, phrases, sentences).

LINE REFERENCES Column 32

1 = Stem directs candidate to a specific line.
2 = Stem direct candidate to 2-4 lines.
3 = Stem directs candidate to a specific paragraph.
4 = Stem does not direct candidate to any specific part of the phrase.

** Hyphenated words, numbers, and Roman numerals were entered as one word.*

SOCIALLY RELEVANT Column 33

A socially relevant sentence is a sentence whose content concerns an aspect or issue of contemporary society that is related to social justice or to political, legal, or economic equality.

1 = Yes	The item is socially relevant.
2 = No	The item is not socially relevant.

SUBJECT CONTENT I Columns 34-35

1 = Humanities	Includes literature, art, music, dance, theater, architecture, religion, philosophy (See also Subject Content II.)
2 = Social Studies	Includes history, sociology, political science, anthropology, general psychology, economics, business.
3 = Biological Sciences -- Traditional	Includes botany, zoology, genetics, Regular microbiology. (See also Subject Content III.)
4 = Biological Sciences -- General	Includes natural history, general animal behavior. (See also Subject Content III.)
5 = Physical Sciences	Includes chemistry, physics, earth science (astronomy, geology). (See also Subject Content III.)
6 = Narrative	Refers to an excerpt from a short story or novel (fiction).
7 = Argumentative Humanities*	
8 = Argumentative* Social Studies	
9 = Argumentative* Science	
10 = Argumentative* Mixed	

SUBJECT CONTENT II Column 36
(A subdivision of the Humanities)

0 = None
1 = Refers to literature only.
2 = Refers to all other humanities.

* *Refers to a passage that takes a stand or tries to persuade*

SUBJECT CONTENT III Column 37
(A subdivision of Biological Sciences and Physical Sciences)

0 = None
1 = Refers to all scientific/technical information or discussion.
2 = Refers to a more general discussion of history, theory, or philosophy of science or to a biographical sketch of a scientist.

ITEM SPECIFICATIONS Columns 38-39

1 = Main Idea	Question asks for the explicit or implicit main idea of the passage.
2 = Main Rhetorical Purpose	Question asks for primary rhetorical purpose (e.g., to present a new point of view).
3 = Best Title	Questions asks for the best or most appropriate title for the passage.
4 = Explicit Information -- Whole Passage	Question asks for identification of supporting details stated in the passage as a whole.
5 = Explicit Information -- Part of Passage	Question asks for identification of supporting details stated in part of the passage.
6 = Inference	Question asks about inferences supported by the passage.
7 = Inference -- Vocabulary	Question asks about the meaning of a word or phrase.
8 = Application	Question asks the candidate to apply information found in the passage to a situation not described in the passage.
9 = Logic	Question asks about the logic of the passage or of an element in the passage.
10 = Organization, Structure, or Rhetorical Devices	Question asks about the structure of the passage or about an element or rhetorical pattern in the passage.
11 = Style	Question asks about the tone or style of all or parts of the passage.

SPECIFIC/NONSPECIFIC Column 40

1 = Stem/Options Specific

An item classified as a 1 would have a stem and options that mention by name, explicitly, actual concepts, details, and/or figures mentioned in the stimulus passage. (Includes close paraphrase or reference)

2 = Stem Nonspecific/ Options Specific

An item classified as a 2 would have a stem that does not mention by name, explicitly, actual concepts, details, and/or figures mentioned in the stimulus passage, but has options that do mention these things. An example of such a stem might be "Which of the following best states the main idea of the passage?"

3 = Stem Specific/ Options Nonspecific

An item classified as a 3 would have a stem, but not options, that mentions actual concepts, details, and/or figures mentioned in the passage.

4 = Stem/Options Nonspecific

An item classified as a 4 would have neither stem not options that mention actual concepts, details, and/or figures mentioned in the passage.

NAMED I - MINORITY IN ITEM Column 41

Refers to the primary individual named or referred to in the passage overall.

0 = No one named or referred to
1 = Black American named or referred to
2 = Hispanic American named or referred to
3 = Native American named or referred to
4 = Asian American named or referred to
5 = Third-world Black named or referred to
6 = Third-world Hispanic named or referred to
7 = Third-world Asian named or referred to
8 = Nonminority ethnic named or referred to
9 = General or unidentifiable

NAMED II - GENDER IN ITEM Column 42

0 = No one named or referred to
1 = Female named or referred to
2 = Male named or referred to
3 = Mixed named or referred to
4 = General or unidentifiable

KIND OF LANGUAGE (Technical/Non-Technical) Column 43

1 = Passage is technical, question is technical or requires very specific knowledge.
2 = Passage is technical, question is general or easily accessible (e.g., tone).
3 = Passage is general, question is general.
4 = Passage is general, question is technical or requires very specific knowledge.
5 = Can't decide.

MEANS OF FINDING ANSWER Column 44
(What appears to be the appropriate or most likely strategy.)

1 = Answer comes to mind after question is read and student then searches for the closest counterpart among options.
2 = Student must look at all options before formulating or identifying answer.
3 = Can't decide.

KEY POSITION -- READING COMPREHENSION Column 45

1 = A
2 = B
3 = C
4 = D
5 = E

STRUCTURAL FOCUS Column 46

1 = Descriptive/Propositional: Any text that focuses primarily on a discussion of one or more ideas, themes, or propositions
2 = Procedural/Narrative: Any text in which progression or process is the key to the structure (e.g., accounts of cyclical or cause and effect phenomena, narratives of character development, accounts of historical consequences)
3 = Argumentative/Persuasive: Any text intended primarily to persuade, convince, anger or, enthuse
4 = Mixture of two of the above (Use this classification sparingly.)

ITEM FORMAT Column 47 (See p. 68)

APPENDIX B
CODING CATEGORIES: MATHEMATICS

SAT VARIABLES IN MATHEMATICS: ALL ITEM TYPES

SAT variables common to all item types:

MINORITY STIMULUS — Column 15

1 = Black	Stimulus refers to Blacks
2 = Hispanic	Stimulus refers to Hispanics
3 = Other Minority	Stimulus refers to other minorities
4 = General	Stimulus refers to people of no specified ethnic origin
5 = Nothing	Stimulus does not refer to people

GENDER REFERENCE IN STIMULUS — Column 16

1 = Female	Stimulus refers to females
2 = Male	Stimulus refers to males
3 = Mixed or unidentifiable	Stimulus refers to both males and females
4 - Neutral	Stimulus does not refer to people

NEGATIVE ITEM (No/Except) — Column 17
(Does not apply to Verbal: Sentence Completions)

1 = Negative Stem	Use of "NOT", "CANNOT", "EXCEPT", "INCORRECT", "FALSE", etc. in stem
2 = Positive Stem	Do not use "NOT", "CANNOT", "EXCEPT", "INCORRECT", "FALSE", etc. in stem

ROMAN NUMERAL FORMAT — Column 18
(Does not apply to Verbal: Sentence Completions)

1 = Roman	Involves Roman numeral format
2 = Non-Roman	Does not involve Roman numeral format

ITEM TYPE Column 19
(Refers to the type of Quantitative item classified)

Example:

	Column A	Column B
1 = Quantitative Comparisons (QC)	s = 6 + 7 + 8 + 9 t = 9 + 8 + 7 + 6	
	1. s + t	4(15)

where the options are:
(A) if the quantity in Column A is greater;
(B) if the quantity in Column B is greater;
(C) if the 2 quantities are equal;
(D) if the relationship cannot be determined from the information given.

2 = Regular Math (5-choice)

Example: If $\frac{y}{x} = -1$, then y + x =

(A) -2 (B) -1 (C) 0 (D) 1 (E) 2

PRIMARY CONTENT AREA Column 20

1 = Arithmetic (ARIT/QCAR)
2 = Algebra (ALGB/QCAL)
3 = Geometry (GEOM/QCAL)
4 = Miscellaneous (MISC/QCMI)

SUB-CONTENT AREA Columns 21, 22

For Arithmetic

10 = Computation
11 = Properties of integers
12 = Properties of rational numbers
13 = Percent
14 = Ration and proportion
15 = Average
16 = Denominate numbers
17 = Tables and charts

For Algebra

20 = Algebraic operations
21 = Word problems
22 = Linear functions
23 = Quadratic functions
24 = Systems of equations & inequalities
26* = Exponents
28 = Series & sequence
29 = Miscellaneous (algebraic averages, permutations, and combinations)

For Geometry

30 = Points, rays, lines in the plane
31 = Angles in the plane (not in triangles, polygons, or circles)
32 = Triangles (not special)
33 = Special triangles
34 = Circles
35 = Polygons (not inscribed or circumscribed)
36**= Polygons (inscribed or circumscribed)
38 = 3-dimensional solids
39 = Coordinate geometry

For Miscellaneous

40 = Structure of the number system
41 = Elementary number system
42 = Sets
43 = Logic
44 = Other SAT (new concepts, probability, geometric perception)
45 = Newly defined operations (contains special symbols/made up definitions)

** Digits 25 and 27 are in the math classification system, but do not apply to the SAT.*

*** Digit 37 is in the math classification system, but does not apply to the SAT.*

Special Topics I Columns 23, 24 Special Topics II Columns 25, 26

00 = Money (same numbers as for Special Topics I)
01 = Time/calendar
02 = Age
03 = Counting problem, probability
04 = Ration/proportion/variation, fractions (but not probability)
05 = Rate (including time and distance)
06 = Liquid measure/weights
07 = Linear measure (perimeter)
08 = Metric system
09 = Numbers/other, includes: temperature, score, letter-arithmetic
10 = Area
11 = Volume
12 = Average (arithmetic mean)
13 = Not applicable
14 = Percent
15 = Angle measures
16 = Endpoint problems

Relationship to Curriculum Column 27

1 = Very textbook-like (clearly in curriculum, standard algorithms apply)

This includes problems that look like the homework problems in a high school Algebra I or Geometry course (as well as straightforward arithmetic examples).

For example, 3HSA024 Section 2, #3, is a standard solution to a linear equation.

3. If $(3n + 6) = 24$, what is the value of n?

(A) 1
(B) 2
(C) 6
(D) 10
(E) 14

Another example, 3HSA024 Section 2, #16, is a standard (but not necessarily easy) ratio problem using some basic geometry.

16. If the degree measures of the angles of a triangle are in a ratio of 2:3:4, what is the degree measure of the greatest angle?

(A) 60
(B) 80
(C) 90
(D) 100
(E) 120

2 = Textbook-like (but less common than above) e.g., new definitions

This refers to problems that are based on standard curriculum, but require unusual steps or combinations of processes.

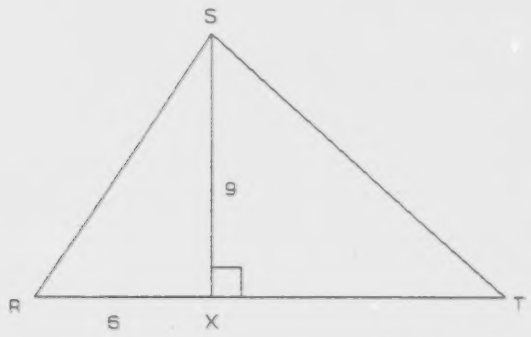
For example, 3HSA024 Section 2, #18, requires the use of standard algebraic techniques, but it is unusual to combine multiplying binomials with solving equations, especially since it asks the student to solve for p^2 rather than p.

18. What is the value of p^2 if $(p + 5)(p - 5) = 24$?

(A) 1 (B) 5 (C) 7 (D) 25 (E) 49

Another example, 3HSA024 Section 2, #17, uses standard school geometry (the formula for the angle of a triangle) but it is unusual in giving the area and asking the student to find part of the length of the base.

Note: Figure not drawn to scale.



17. If the area of $\triangle RST$ above is 90, what is the length of XT?

(A) 4 (B) 7 (C) 10 (D) 14 (E) 20

A third example, 3HSA024 Section 2, #25, looks untextbook-like because of the odd symbolism, N*, but the concepts involved, such as adding up the whole numbers between 1 and N and considering whether the sum is odd or even, are standard concepts of arithmetic.

25. For any positive whole number N, let N* equal the sum of all whole numbers between 1 and N, inclusive; for example:
$4^* = 1 + 2 + 3 + 4 = 10$.

Which of the following statements must be true?

I. 20^* is an odd number.
II. If P is a positive odd whole number, then P^* is odd.
III. If R is a positive whole number, then $(R + 1)^* - R^*$ is equal to $(R + 1)$.

(A) None (B) I only (C) II only
(D) III only (E) II and III

All problems involving newly-defined operations were classified the same way for column 27.

3 = Not textbook-like (depends on practical experiences outside of school or has novel application)

This refers to problems that are easier for students who can observe mathematical patterns.

For example, 3HSA024 Section 2, #5, is easier to do if you have observed how scores are recorded on scoreboards.

5. The scoreboard below shows the end-of-quarter cumulative scores for two teams.

	END OF QUARTER			
	1	2	3	4
VISITORS	14	22	30	46
HOME TEAM	8	23	38	46

What was the greatest number of points scored by either team in a single quarter?

(A) 14 (B) 15 (C) 16 (D) 17 (E) 18

Another example, 3HSA024 Section 5, #20, requires students to think about the patterns of digits in an addition problem.

Column A | Column B

```
  99
+ YY
 XY6
```

X and Y represent different digits in the correctly worked addition problem above.

20. Y | 5

All letter-addition problems were classified the same for column 27.

A third example, 3HSA035 Section 2, #9, requires students to recognize and use patterns in a specified set of numbers.

```
1
1 3
1 3 5
1 3 5 7
```

9. In the figure above, the first row contains the first of the positive odd integers, the second row contains the first two of the positive odd integers, the third row contains the first three of the positive odd integers, and so on. If the figure is continued in this fashion, what is the sum of the integers in the tenth row?

(A) 1,024 (B) 512 (C) 256 (D) 100 (E) 81

Generally, problems involving pattern recognition were classified the same way for column 27.

Ability Level Column 28

0 = Recall factual knowledge
1 = Perform mathematical manipulations
2 = Solve routine problems
3 = Demonstrate comprehension of mathematical ideas & concepts
4 = Solve non-routine problems requiring insight or ingenuity
5 = Apply "higher" mental processes to mathematics

Item Attributes (Regular Math only) Column 29

QSUFF - "It cannot be determined from the information given" is option E.

1 = QSUFF
2 = NON QSUFF
3 = Does not apply (QC problem)

Max/Min Column 30 Regular Math or QC

1 = Attribute present
2 = Attribute absent
3 = Does not apply

Must/Could Column 31 (Regular Math or QC)

1 = *Must* appears in stem
2 = *Could* appears in stem
3 = Does not apply

Type of solution Column 32
*See notes below on Roman Numeral, QC.

1 = Computed solution (choices: numbers, or sets, or points)
2 = General solution (formula, literal expression, conceptual choices, or inequality)
3 = Does not apply

Role of options Column 33

1 = Solution independent of options
2 = Solution requires examination of options
3 = Does not apply (QC)

Order of options Column 34

1 = Options are listed least to greatest
2 = Options are listed greatest to least
3 = Does not apply or mixed, including QC, Roman Numeral type

** Type of solution*
Roman Numeral
If the Roman Numeral choices (I, II, III) are computed choices (numbers, etc.), code 1. If they are general expressions, etc.), code 2.

QC
If you need to find exact values, code 1. If you use general information to compare the quantities, code 2.

**Spatial Factor Column 35

1 = Primary spatial component
2 = No figure shown, but drawing or sketch would help
3 = Possible spatial factor (e.g., figure in which extra lines needed or with fairly complex visual field
4 = Estimation helpful in eliminating at least two of the options
5 = Probably not a spatial factor
6 = Ordinary geometry

Scale of figure Column 36
(Refers to triangles, squares, etc., not charts, tables, or graphs)

1 = Figure not drawn to scale (note is present)
2 = Figure is drawn to scale
3 = No figure, or not applicable

READING DIFFICULTY - I Column 37

1 = Difficult Items with compound sentences are/or large numbers of words perhaps requiring logic to sort out the meaning. Items which require careful reading.

Example: Worker W produces n units in 5 hours. Workers V and W, working independently but at the same time, produce n units in 2 hours. How long would it take V alone to produce n units?

2 = Medium Items with less verbiage; contain a simple word, phrase, or short sentences. Meaning is readily clear.

Example: A certain photocopying machine can make 10 copies every 4 seconds. At this rate, how many copies can the machine make in 6 minutes?

** *Spatial factor: Code 4 if it applies, even if some other code also applies.*

3 = Easy Items which do not contain words or items which contain only a few (at most) standard words, such as (A) if ———, then ———; (B) ——— and ———; (C) if ——— and ———, then ———; (D) in the figure above.

Example: (A) If $\frac{y}{x} = -1$, then $y + x =$

(B) $x = 9$ and $y = 3$

(C) If $2x + 3y = 15$, and $y = 1$, then $2x =$

(D) In the figure above, x = (without any further explanation given, other than the figure. If a more detailed explanation is given in the stem, the reading difficulty would be a 2.)

CONCRETE/ABSTRACT Column 38

1 = Concrete Questions that are real-life word problems.

Example: A supervisor was paid for her travel expenses at the rate of $0.20 per mile. If she received $14.40, for how many miles was she paid?

2 = Abstract Questions that do not involve real-life settings.

Example: What is the sum of the areas of two squares with sides of lengths 1 and 3, respectively?

MULTIPLE CATEGORIES Column 39

1 = Problems that can be solved using arithmetic only
2 = Problems that can be solved using arithmetic and/or algebra, including primarily algebra
3 = Problems that can be solved using arithmetic and/or geometry, including primarily geometry
4 = Problems that can be solved using algebra and/or geometry
5 = Problems that can be solved using logic

STIMULUS FORMAT I: PICTURES Column 40

1 = Figure Picture does not have a coordinate system (has a triangle, square, rectangle, etc.)
2 = Graph Picture has a coordinate system or is a line, bar, or circle graph
3 = Table Picture has data presented in rows and columns, including magic squares, times tables, letter arithmetic
4 = None Pictures are not included in the stimulus.
5 = Combination of 1, 2, or 3
6 = Number line
7 = Venn diagram
8 = Picture Actual sketches or drawings of objects (trees, logs, buildings, etc.)

STIMULUS FORMAT II: VARIABLES (Include QC) Column 41

1 = Variables are present in the options only.
2 = Variables are present in the stem or stimulus only.
3 = Variables are present in the options and the stem or the stimulus.
4 = Variables are not present.

KEY - QUANTITATIVE COMPARISON PROBLEMS ONLY Column 42

(See example given under ITEM TYPE: Quantitative Comparisons for a complete description of options A through D.)

1 = A
2 = B
3 = C
4 = D
5 = Not applicable (Regular Math)

UNDERLINING IN STEM OR STIMULUS Column 43

1 = Attribute present (but exclude "Note: Figure not drawn to scale.")
2 = Attribute not present

LENGTH OF STIMULUS Columns 44-46

Actual word counts Does not include symbols, numbers, single letters, Roman numerals, or formulae

LENGTH OF STEM Columns 47-49

Actual word counts Does not include symbols, numbers, single letters, Roman numerals, or formulae

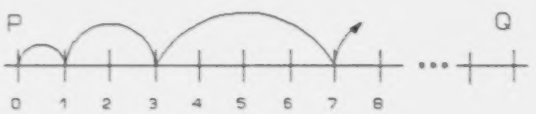
KEY -- REGULAR MATH Column 50

1 = A
2 = B
3 = C
4 = D
5 = E
6 = Not applicable (QC)

READING DIFFICULTY - II Column 51

1. Difficult Complicated grammar, many words, complicated logical connections, the use within an item of the same word or concept but with more than one role in the item, changes in the order in which information is given, shifts in point of view, and unexpected completion of sentences are all factors that contribute to item difficulty. The classification "difficult" (difficult to read, that is) is based on the cumulative effect of these factors.

For example, 3HSA034 Section 2, #15 contains many words and uses "hop" as both a verb and a noun. The second sentence is complicated further by the introductory prepositional phrase and the long comparison.



15. A flea hops along a number line starting at point *P*, as shown above. With each hop after the first, it hops twice as far as it did on the preceding hop. If the flea stops at point *Q*, which of the following numbers could NOT be the number at *Q*?

(A) 15 (B) 31 (C) 47 (D) 63 (E) 127

Another example, 3HSA034 Section 5, #29, speaks of numbers of numbers, "three 3's and one 2," and numbers of ways of expressing a number. Although the item uses the word "number" only once, the repeated use of the concept of number makes it important for the student to read this carefully.

29. The number 11 can be expressed as the sum of 2's and 3's in two ways, that is 3 + 3 +3 + 2 (three 3's and one 2) or 3 + 2 + 2 + 2 + 2 (one 3 and four 2's). In how many ways can 17 be expressed as the sum of 2's and 3's?

(A) One
(B) Two
(C) Three
(D) Four
(E) Five

A third example is 3HSA034 Section 2, #22, in which the first sentence is long and complicated. Not only does this sentence contain a subordinate noun clause, but this noun clause has a complicated subject, "the number of chirps a cricket makes in 15 seconds" followed by the verb "is". The reader is led to expect a short completion to this clause, such as a number. A number, "40", does follow "is", but, as the reader reads on, it becomes clear that "40" is not the predicate subject; "40 less than the temperature in degrees Fahrenheit" is the predicate subject. (This is an example of an unexpected completion of a sentence.)

22. It is estimated that the number of chirps a cricket makes in 15 seconds is 40 less than the temperature in degrees Fahrenheit. At this rate, if a cricket chirps n times in x minutes, what is the temperature in degrees Fahrenheit?

(A) $\frac{n}{x} + 40$

(B) $\frac{n + 40}{x}$

(C) $\frac{4n + 40}{x}$

(D) $\frac{n}{4x} + 40$

(E) $\frac{n + 40}{4x}$

Another example, 3HSA024 Section 2, #24, shows a shift in point of view in the second sentence, which addresses the reader-as-problem-solver, as compared with the first sentence, which is straight exposition.

24. Kate mailed a letter that weighed w ounces. Assume w is an integer greater than 1. If the postage rate was 18 cents for the first ounce and 6 cents for each additional ounce, which of the following gives the postage cost, in cents, for the letter?

(A) $w + 24$
(B) $6w + 12$
(C) $6w + 17$
(D) $6w + 18$
(E) $18w + 6$

A shift in the order of presentation of information can be seen in 3HSA034 Section 2, #3, in which the first sentence mentions 7:00 a.m. followed by 4:00 p.m., while the second sentence starts with 4:00 p.m.

3. Between 7:00 a.m. and 4:00 p.m. on a certain day, the temperature rose 28 degrees. If the temperature at 4:00 p.m. was 20 degrees above zero, what was the temperature at 7:00 a.m.?

(A) 24° below zero
(B) 20° below zero
(C) 8° below zero
(D) 8° above zero
(E) 48° above zero

Newly-defined operations with a highly verbal component, such as K-3GSA026 Section 2, #21, are generally classified as difficult to read.

For any positive integer k, $(k)\#$ represents the greatest odd numbers that divides k: for example, $(36)\# = 9$

21. $(2^3 \cdot 5^2)\#$ $(2^4 \cdot 3 \cdot 5)\#$

The use of logical connectors such as "any", "either", "neither", and "both" often contributes to item difficulty. An example of this is 3HSA024 Section 2, #5, in which the student needs to find the greatest element of a certain set, and the definition of that set depends partly on the phrase "scored by either team."

5. The scoreboard below shows the end-of-quarter cumulative scores for two teams.

	END OF QUARTER			
	1	2	3	4
VISITORS	14	22	30	46
HOME TEAM	8	23	38	46

What was the greatest number of points scored by either team in a single quarter?

(A) 14 (B) 15 (C) 16 (D) 17 (E) 18

Item #10 of 3HSA024 Section 2 tests verbal logic and thus is classified as difficult to read.

10. On a certain island it is known that any person who has brown hair does not eat fish. Which of the following statements must also be true of people on the island?

I. Any person who eats fish does not have brown hair.
II. All persons who do not eat fish have brown hair.
III. Any person who does not have brown hair eats fish.

(A) I only (B) II only (C) I and III only
(D) II and III only (E) I, II, and III

2. Easy or Medium

Examples include items with very few words, such as 3HSA034 Section 2, #4,

4. If $\frac{x}{y} = 1$, then $3x - 3y =$
(A) -1 (B) 0 (C) ⅓ (D) 1 (E) 3

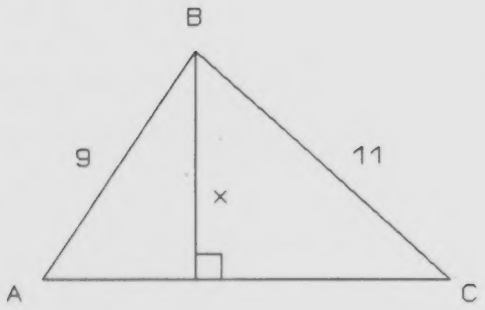
as well as items with more words, provided that the sentences are not very long or complicated. Two examples are 3HSA034 section 2, #6

6. Three people together buy a pizza for $4.25 and leave a tip equal to 20 percent of the price of the pizza. If they share the cost of the pizza and the tip equally, how much does each person pay?

(A) $0.85
(B) $1.10
(C) $1.40
(D) $1.70
(E) $2.55

and 3HSA034 Section 5, #28.

Note: Figure not drawn to scale.



28. If the perimeter of $\triangle ABC$ above is 32, what is the area of $\triangle ABC$ in terms of x?

(A) 12x (B) 10x (C) 9x (D) 6x (E) 5x

Items involving newly defined operations are included here if the definition is algebraic rather than primarily verbal. An example is 3HSA034 Section 5, #30.

30. If the operation ⊗ is defined for all numbers a and b by the equation $a \otimes b = \frac{a + b}{2}$, then

$2 \otimes (4 \otimes 8) =$

(A) 4 (B) 4½ (C) 5½ (D) 6 (E) 7

Roman numeral format may contribute to reading difficulty, but not all such items are classified as difficult. For example, 3HSA034 section 5, #35 is a mathematically difficult item in Roman numeral format, but it is not difficult to read.

35. If *a*, *b*, *c*, *and* *e* are positive integers, and if the expression $bc(a + e)$ is an odd number, which of the following numbers could be <u>even</u>?

I. b
II. b + c
III. a + e

(A) None (B) I only (C) II only
(D) I and II (E) II and III

Although logical connectors such as "either" and "both" may contribute to reading difficulty, the presence of such a word in the item does not necessarily make the item difficult to read. For example, 3HSA024 Section 5, #22, uses "both" but is not classified as difficult to read.

The area of circle *C* and square *S* are both equal to 16π.

22. The radius of circle *C* The length of a side of square *S*.

<u>ITEM FORMAT</u> Column 52
(Refers to the form used to set up the item or the way the item appears in the test.)

1 = Type 1
(each choice on separate line or arranged vertically)

CHOIR: SINGER:
(A) election: voter
(B) anthology: poet
(C) cast: actor
(D) orchestra: composer
(E) convention: speaker

2 = Type 2
(choices run together or arranged horizontally on two lines or more)

WATER: SWIM:: (A) grass: grow
(B) knot: tie (C) plan: implement
(D) flood: damage (E) snow: ski

3 = Type 3
(choices run together or arranged horizontally on one line)

A - B - C - D - E -

4 = Not Applicable (QC)

APPENDIX C

CODING CATEGORIES: TEST OF STANDARD WRITTEN ENGLISH (TSWE)
SAT VARIABLES COMMON TO ALL TSWE ITEM TYPES

MINORITY STIMULUS Column 18

1 = Black Americans	Stimulus refers to Black Americans
2 = Hispanic Americans	Stimulus refers to Hispanic Americans
3 = Native Americans	Stimulus refers to Native Americans
4 = Asian Americans	Stimulus refers to Asian Americans
5 = Third-World Blacks	Stimulus refers to Third-World Blacks
6 = Third-World Hispanic	Stimulus refers to Third-World Hispanics
7 = Third-World Asian	Stimulus refers to Third-World Asians
8 = Nonminority Ethnic	Stimulus refers to other ethnic groups
9 = General	Stimulus refers to people of no specified ethnic origin
0 = Nothing	Stimulus does not refer to people

GENDER REFERENCE IN STIMULUS Column 19

1 = Female	Stimulus refers to females only
2 = Male	Stimulus refers to males only
3 = Mixed	People referred to in stimulus are unidentified as to whether they are male or female (such as teachers, they, we, you, students)
5 = Neutral	Stimulus does not refer to people

NEGATIVE ITEM (No/Except) Column 20

1 = Negative Stem	Use of "NOT", "CANNOT", "EXCEPT", "LEAST", "INCORRECT", "FALSE" etc., in stem
2 = Positive stem	Do not use "NOT", etc., in stem

ROMAN NUMERAL FORMAT Column 21

1 = Roman	Involves Roman numeral format
2 = Non-Roman	Does not involve Roman numeral format

ITEM FORMAT Column 22

1 = Usage
2 = Sentence Correction

EMOTIVE QUALITY Column 23

0 = Sentence refers to neutral or pleasant subject matter.

1 = Sentence refers to strongly upsetting subject matter (e.g., evil, fire, flood, nuclear war.) NOTE: This list will be expanded during coding. This does not include argumentative or inflammatory subject coding; rather it refers to questions which have a negative impact or an overall tone of a depressing nature. The word strongly is the clue here.

2 = Can't Decide (NOTE: Use can't decide as a flag or signal that the coding descriptions need clarifying or that another opinion is needed. Ultimately, all items should fit into the coding categories.)

SAT VARIABLES STUDIED IN TSWE

(See also "Variables Common to All TSWE Item Types").

SPECIFICATIONS Columns 24-25

1 = Subject-verb agreement with interrupting phrase
2 = Subject-verb agreement after expletive
3 = Subject-verb agreement with inverted structure
4 = Subject-verb agreement neither/nor, either/or
5 = Tense sequence
6 = Word clue to tense
7 = Verb form
8 = Nonidiomatic connective
9 = Wrong relative pronoun
10 = Logical agreement
11 = Logical comparison
12 = Adjective/adverb confusion
13 = Double negative
14 = Comparison of adjectives
15 = Pronoun case
16 = Pronoun shift
17 = Unclear pronoun reference
18 = Lack of pronoun agreement
19 = Diction (common errors)
20 = Idiomatic preposition
21 = Idiomatic structure
22 = Idiomatic infinitive/participle
23 = No error
24 = Parallelism
25 = Sentence fragment
26 = Comma splice
27 = Improper subordination
28 = Improper coordination

Specifications, con't.

29 = Dangling modifier
30 = Redundancy/economy/conciseness/clarity
31 = Vague pronoun reference (this, it)
32 = Illogical comparison
33 = Subject shift
34 = Fused sentence
35 = Active passive shift
36 = Misplaced modifier

CONTENT OF SENTENCE Column 26

1 = Arts
2 = Social Science
3 = Science
4 = Public life
5 = Student relevant
6 = Everyday activities

UNDERLINE POSITION Column 27 (Sentence Correction Only)

1 = Includes first word
2 = Does not include first or last word
3 = Includes last word
4 = Includes entire sentence
5 = Not applicable (Refers to Usage item type)

LENGTH OF STEM Columns 28-30

Enter total number of words in sentence (includes "no error" in Usage).

LENGTH OF OPTIONS Columns 31-33
(A through D or E)

For Usage, enter total number of underlined words in sentence (Includes "non error").

For Sentence Correction, enter the total word count for Options A through E.

SENTENCE STRUCTURE Column 34

1 = Simple sentence
2 = Compound sentence
3 = Complex sentence -- 1 dependent clause
4 = Complex sentence -- 2 or more dependent clauses
5 = Compound - complex sentence

KEY — Column 35

1 = Error
2 = No error — The A key in Sentence Correction and the E key in Usage are the no-error options.

CLUE TO KEY — (Usage item type only) Column 36

1 = Occurs before the first underline
2 = Occurs after first, but before keyed underline (in B-, C-, D-key sentences)
3 = Occurs after keyed underline
4 = Occurs within keyed underline (e.g., hardly no)
0 = No single clue to key in sentence

NAMED I - MINORITY ITEM — Column 37
(Refers to a person or persons specifically named such as Sarah Jones, Abraham Lincoln, etc.)

0 = No one named or referred to
1 = Black American named or referred to
2 = Hispanic American named or referred to
3 = Native American named or referred to
4 = Asian American named or referred to
5 = Third-world Black named or referred to
6 = Third-world Hispanic named or referred to
7 - Third-world Asian or named referred to
8 = Nonminority ethnic named or referred to
9 = General or unidentifiable

NAMED II - GENDER IN ITEM — Column 38
(Refers to someone specifically named.)

0 = No one named or referred to
1 = Female named or referred to
2 = Male named or referred to
3 = Mixed named or referred to
4 = General or unidentifiable

CONTROVERSIAL SUBJECT MATTER — Column 39

Refers to subject matter that could be controversial or inflammatory in nature, such as women's rights, police violence, pollution, use of pesticides, etc. Include topics which are sociopolitical in nature.

1 = Yes — Attribute is present.
2 = No — Attribute is not present.

TSWE -- SPECIFICATIONS -- RECOMBINING

	Specifications
1. Subject/Verb Agreement	1, 2, 3, 4
2. Tense	5, 6
3. Logical Comparison & Agreement	10, 11, 32
4. Pronouns	15, 17, 18, 31
5. Idiom	8, 20, 21, 22
6. Diction	19
7. Usage Conventions	7, 9, 12, 13, 14
8. No Error	23
9. Shift	16, 33, 35
10. Sentence Boundaries	25, 26, 34
11. Sentence Joining	27, 28
12. Dangling or Misplaced Modifiers	29, 36
13. Parallelism	24
14. Clarity/Economy	30

Table 1
Summary of SAT and TSWE Test Form Administration Dates and Sample Sizes

Form	Administration Date	Gender		Ethnic/Racial Background				
		Male	Female	White	Asian	Black	Hispanic	Am. Indian
SAT								
1	January 1983	6,329	6,712	9,650	611	1,738	498	79
2	January 1984	25,986	27,158	39,960	2,154	7,128	1,864	321
3	October 1984	14,529	17,822	25,120	2,367	2,392	1,509	132
4	March 1985	31,588	33,038	53,781	2,515	5,105	1,420	269
5	March 1985	28,553	29,993	48,861	2,241	4,578	1,252	236
6	December 1985	74,243	83,945	102,442	6,185	19,243	7,081	2,003
Total		181,228	198,668	279,814	16,073	40,184	13,624	3,040
TSWE								
1	March 1985	60,141	63,031	102,642	4,756	9,683	2,672	505
2	December 1985	14,959	16,968	24,450	1,226	3,843	1,390	377
3	December 1985	14,724	16,843	24,096	1,239	3,791	1,413	402
4	December 1985	16,831	19,096	26,799	1,432	4,667	1,740	484
5	December 1985	13,933	15,900	22,852	1,184	3,513	1,320	373
6	December 1985	13,796	15,138	22,245	1,104	3,429	1,218	367
Total		134,384	146,976	223,084	10,941	28,926	9,753	2,508

Table 2
Summary of Mean Formula Score Differences* Between Reference and Focal Groups
in Standard Deviation Units

Item Categories	Gender		White/Black	White/Hispanic
	Male/Female	White/Asian	White/Black	Mean S.D.
SAT - Verbal				
Form 1	-.14	-.12	-.97	-.64.
Form 2	-.12	-.16	-.96	-.62
Form 3	-.14	-.02	-.92	-.54
Form 4	-.15	+.02	-1.09	-.60
Form 5	-.14	+.07	-1.04	-.65
Form 6	-.11	-.10	-.87	-.56
SAT - Mathematics				
Form 1	-.38	+.31	-.99	-.59
Form 2	-.44	+.24	-.99	-.59
Form 3	-.46	+.39	-.95	-.49
Form 4	-.40	+.36	-1.06	-.59
Form 5	-.38	+.32	-1.09	-.63
Form 6	-.42	+.25	-.92	-.55

* Calculated by the formula: $\frac{\text{(Mean Formula Score Focal Examinees - Mean Formula Score Reference Examinees)}}{\text{Total Group Standard Deviation}}$. Thus, a negative value indicates that the Reference Group examinees (i.e., Males and Whites) perform better than the focal group examinees, and a positive value indicates that the focal group examinees perform better than the reference group examinees.

BEST COPY AVAILABLE

Table 3
Average DIF* Values for Reference/Focal Comparison Groups

Item Type	Male/Female		White/Asian		White/Black		White/Hispanic	
	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
SAT - Verbal								
Antonyms (150 items)	-.07	.78	-.07	.66	.13	.60	.02	.53
Analogies (120 items)	-.07	.71	-.06	.58	-.15	.62	-.19	.57
Sentence Completions (90 items)	.04	.62	.02	.54	.04	.40	-.17	.56
Reading Comprehension (150 items)	.12	.55	-.04	.35	.06	.34	.16	.29
SAT - Mathematics								
Quantitative Comparison (120 items)	-.00	.50	.02	.53	.05	.42	.01	.29
Regular Mathematics (240 items)	-.02	.52	-.02	.68	.01	.47	.03	.35
Total (360 items)	-.01	.51	-.01	.64	.03	.45	.02	.32
TSWE								
Usage (210 items)	-.01	.39	-.09	.81	.01	.52	.01	.46
Sentence Correction (90 items)	.08	.43	.18	.55	.08	.44	.08	.37
Total (300 items)	.02	.41	-.01	.75	.03	.50	.03	.44

* Negative values indicate that for gender comparisons, male examinees perform relatively better than female examinees with comparable test scores. Similarly, for ethnic comparisons, negative values indicate that White examinees perform relatively better than focal group examinees with comparable test scores.

Table 4
Number and Percent of SAT Items by Section and Reference/Focal Group
Comparison that Evidence DIF Greater Than 1.0 or Less Than -1.0

Item Type		Items with DIF<-1.0*				Items with DIF>1.0*			
		Gender	Ethnic/Racial Background			Gender	Ethnic/Racial Background		
		M/F	W/Asian	W/B	W/H	M/F	W/Asian	W/B	W/H
SAT-Verbal									
Analogies	N	10	8	17	10	7	1	0	1
(120 items)	%	8.3	6.7	14.2	8.3	5.8	.8	0.0	.8
Antonyms	N	14	14	6	7	6	4	5	6
(150 items)	%	9.3	9.3	4.0	4.7	4.0	2.7	3.3	4.0
Sentence Completions	N	5	4	2	2	2	3	1	0
(90 items)	%	5.6	4.4	2.2	2.2	2.2	3.3	1.1	0.0
Reading Comprehension	N	7	2	1	0	5	0	1	0
(150 items)	%	4.7	1.3	.7	0.0	3.3	0.0	.7	0.0
Total	N	36	28	26	19	20	8	7	7
(510 items)	%	7.1	5.5	5.1	3.7	3.9	1.6	1.4	1.4
SAT-Mathematics									
Quantitative Comparison	N	2	4	2	0	3	3	1	0
(120 items)	%	1.7	3.3	1.7	0	2.5	2.5	.8	0
Regular Mathematics	N	8	20	6	2	6	13	2	0
(240 items)	%	3.3	8.4	2.5	.8	2.5	5.4	.8	0.0
Total	N	10	24	8	2	9	16	3	0
(360 items)	%	2.8	6.7	2.2	.6	2.5	4.4	.8	0.0
TSWE									
Usage	N	1	21	5	4	1	24	4	3
(210 items)	%	.5	10.0	2.4	1.9	.5	11.4	1.9	1.4
Sentence Correction	N	1	1	0	0	0	6	1	1
(90 items)	%	1.1	1.1	0.0	0.0	0.0	6.7	1.1	1.1
Total	N	2	22	5	4	1	30	5	4
(300 items)	%	.7	7.3	1.7	1.3	.3	8.3	2.7	1.3

* Items with DIF<-1.0 indicate that reference group examinees perform relatively better than focal group examinees with comparable test scores; items with DIF>1.0 indicate the reverse.

Table 5
Number and Percent of SAT Items by Section and Reference/Focal Group Comparison that Evidence DIF Greater Than 1.5 or Less Than -1.5

Item Categories		Items with DIF<-1.5*				Items with DIF 1.5*			
		Gender	Ethnic/Racial Background			Gender	Ethnic/Racial Background		
		M/F	W/Asian	W/B	W/H	M/F	W/Asian	W/B	W/H
SAT-Verbal									
Analogies	N	5	2	4	5	1	0	0	0
(120 items)	%	4.2	1.7	3.3	4.2	.8	0.0	0.0	0.0
Antonyms	N	6	2	4	2	2	1	2	0
(150 items)	%	4.0	1.3	2.7	1.3	1.3	.7	1.3	0.0
Sentence Completions	N	1	1	0	0	0	0	0	0
(90 items)	%	1.1	1.1	0.0	0.0	0.0	0.0	0.0	0.0
Reading Comprehension	N	1	0	0	0	1	0	0	0
(150 items)	%	.7	0.0	0.0	0.0	.7	0.0	0.0	0.0
Total	N	13	5	8	7	4	1	2	0
(510 items)	%	2.5	1.0	1.6	1.4	.8	.2	.4	0.0
SAT-Mathematics									
Quantitative Comparison	N	0	0	0	0	0	0	0	0
(120 items)	%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Regular Mathematics	N	1	8	1	0	0	2	0	0
(240 items)	%	.4	3.3	.4	0.0	0.0	.8	0.0	0.0
Total	N	1	8	1	0	0	2	0	0
(360 items)	%	.3	2.2	.3	0.0	0.0	.6	0.0	0.0
TSWE									
Usage	N	0	8	2	2	0	4	0	1
(210 items)	%	0.0	3.8	1.0	1.0	0.0	1.9	0.0	.5
Sentence Correction	N	0	0	0	0	0	0	0	1
(90 items)	%	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.1
Total	N	0	8	2	2	0	4	0	2
(300 items)	%	0.0	2.7	.7	.7	0.0	1.3	0.0	.7

* Items with DIF<-1.5 indicate that reference group examinees perform relatively better than focal group examinees with comparable test scores; items with DIF>1.5 indicate the reverse.

Table 6
Means and Standard Deviations for Item Categories on the Analogies Section of the SAT-V: Points Tested

Item Categories	Number of Items	Comparison							
		Male/Female		White/Asian		White/Black		White/Hispanic	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Semantic Relationship									
Class Inclusion	2	-.15	1.27	.29	.52	-.42	.07	-.36	.12
Part-Whole: Pos.	10	.12	.41	-.33	.45	-.61	.50	-.62	.42
Similar: Degree	8	-.10	.87	-.37	.64	-.35	.76	-.34	.62
Similar: Other	8	-.32	.53	-.18	.61	-.47	.72	-.52	.64
Attribute	24	-.09	.69	.13	.52	.01	.55	-.07	.52
Contrast	8	-.16	.77	.12	.40	.21	.38	.29	.43
Non-Attribute	11	.30	.24	.05	.55	.00	.57	-.15	.38
Case Relation	15	-.36	.73	-.15	.57	-.16	.52	-.35	.50
Cause/Purpose	17	-.07	.49	.01	.53	.09	.61	.11	.54
Space/Time	11	-.16	1.05	-.24	.74	-.51	.63	-.58	.39
Representation	5	.47	.43	.05	.42	.11	.25	.28	.53
Other	1	-.27	-	.14	-	-.37	-	-.03	-
F		.93		1.01		2.07**		3.31***	

Note: Total number of SAT-V Analogies items = 120.

*p < .10. **p < .05. ***p < .01.

Table 7
Means and Standard Deviations for Item Categories on the Antonyms Section of the SAT-V: Points Tested

Item Categories	Number of Items	Comparison							
		Male/Female		White/Asian		White/Black		White/Hispanic	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Kinds of Antonyms									
General Definition	104	- .11	- .83	- .06	.66	.09	.63	- .03	.52
Fine Distinction	46	.02	- .63	- .08	.64	.23	.51	.14	.53
F		.91		.01		1.93		3.49*	
Parts of Speech									
Verb	49	- .19	.80	- .22	.68	.19	.65	- .11	.51
Noun	41	.02	.70	.12	.52	.13	.66	.08	.55
Adjective or Adverb	60	- .04	.79	- .07	.69	.08	.50	.09	.52
F		.88		3.02**		.44		2.24	

Note: Total number of SAT-V Antonyms items = 150.

*p < .10. **p < .05. ***p < .01.

Table 8
Means and Standard Deviations for Item Categories on the Reading Comprehension Section of the SAT-V: Points Tested

Item Categories	Number of Items	Comparison							
		Male/Female		White/Asian		White/Black		White/Hispanic	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Item Specification									
Main Idea	11	.11	.56	.10	.33	.06	.19	.24	.28
Rhetorical	10	.13	.36	.15	.19	.23	.27	.20	.15
Best Title	5	-.12	.46	.01	.15	.39	.31	.30	.26
Whole Passage	17	.22	.60	-.05	.33	.08	.32	.16	.24
Part Passage	25	.03	.59	.12	.33	.11	.44	.24	.31
Inference	37	.10	.54	-.17	.37	.03	.32	.10	.30
Inference-Vocabulary	3	.67	.17	-.27	.09	.41	.08	.49	.14
Application	12	-.09	.62	-.06	.33	-.02	.27	.21	.33
Logic	6	.16	.48	-.09	.33	-.09	.41	.13	.28
Rhetorical Device	12	.44	.27	-.17	.38	-.14	.27	.00	.18
Style	12	.01	.54	-.10	.31	-.04	.25	.03	.19
F		1.17		2.02**		1.89**		1.74*	

Note: Total number of SAT-V Reading Comprehension items = 150.

*p < .10. **p < .05. ***p < .01.

Table 9
Means and Standard Deviations for Item Categories on the SAT-M: Points Tested

Item Categories	Number of Items	Comparison							
		Male/Female		White/Asian		White/Black		White/Hispanic	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Item Type									
Quantitative Comparison	120	- .00	.50	.02	.53	.05	.42	.01	.29
Regular Mathematics	240	- .02	.52	- .02	.68	.01	.47	.03	.35
F		.06		.43		.70		.25	
Primary Content									
Arithmetic	108	- .09	.52	- .18	.54	.03	.48	- .05	.30
Algebra	103	.03	.51	- .03	.70	.02	.48	.01	.34
Geometry	98	- .13	.48	.33	.61	- .01	.42	.15	.28
Miscellaneous	51	.28	.41	- .24	.66	.11	.40	- .07	.35
F		8.92***		16.17***		.81		9.13***	
Multiple Categories									
Arithmetic Only	124	- .08	.54	- .32	.62	- .08	.49	- .09	.30
Arithmetic/Algebra	133	.14	.47	.07	.57	.17	.41	.04	.32
Arithmetic/Geometry	72	- .19	.48	.30	.53	- .05	.43	.13	.30
Algebra/Geometry	29	.01	.42	.31	.45	.08	.36	.19	.25
Logic	2	.60	.24	- 1.55	1.04	- .63	.08	- 1.05	.35
F		7.05***		20.04***		7.28***		15.11***	
Presence of Variable									
Stem/Stimulus	101	.13	.43	.09	.55	.10	.37	.08	.29
Stem/Stim/Opts.	93	.11	.44	.15	.50	.18	.41	.10	.32
Not Present	166	-.16	.55	-.15	.72	-.10	.48	-.06	.34
F		14.26***		8.86***		14.34***		10.19***	
Ability Level									
Factual Knowledge	1	.66	-	- .21	-	.84	-	- .13	-.15
Math Manipulation	12	.31	.53	.30	.69	.24	.50	.12	.37
Routine Problem	38	.13	.54	.17	.63	- .01	.47	.03	.32
Math Concept	197	- .01	.53	.03	.64	.01	.47	.03	.31
Non-routine	73	- .09	.47	- .08	.54	.01	.38	.03	.33
Higher Mental	38	- .14	.34	- .35	.62	.05	.43	- .11	.37
F		2.79**		3.86***		1.28		1.55	
Type of Solution									
Computed Solution	256	- .04	.51	- .04	.66	- .00	.45	- .00	.32
General Solution	104	.07	.49	.08	.58	.10	.46	.07	.35
F		3.94**		2.48		3.56*		3.50*	
Special Topics									
Money	13	- .23	.32	- .13	.59	.13	.54	.04	.34
Time/Calendar	14	- .51	.51	- .57	.47	- .38	.40	- .17	.30
Age	1	- .84	-	- 1.19	-	- .20	-	- .09	-
Counting	4	.13	.49	- .69	.58	- .05	.61	- .40	.10
Fractions	48	.05	.47	.14	.47	.15	.40	.07	.22
Rate	8	- .73	.26	- .71	.56	- .48	.48	- .18	.32
Liquid Measure	3	- .34	.32	- .36	.47	.25	.52	- .07	.44
Linear Measure	69	- .17	.50	.27	.59	- .08	.46	.12	.31
Numbers	11	- .08	.68	- .95	.74	- .39	.58	- .32	.37
Area	7	- .22	.46	.26	.39	- .00	.55	.07	.37
Average	8	- .37	.42	- .34	.43	.05	.40	- .29	.17
Percent	11	- .34	.52	- .28	.70	.03	.45	- .08	.34
Angle	21	- .09	.45	.39	.31	.01	.35	.22	.24
Endpoint	2	.02	.07	- .02	.28	.23	.15	- .10	.11
Not Applicable	140	.25	.39	- .01	.60	.13	.38	.02	.33
F		8.12***		7.35***		3.66***		3.98***	

Note: Total number of SAT-M items = 360. *p < .10. **p < .05. ***p < .01.

Table 10
Means and Standard Deviations for Item Categories on the SAT-M:
Points Tested (Subcontent Areas)

Item Categories	Number of Items	Comparison							
		Male/Female		White/Asian		White/Black		White/Hispanic	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Arithmetic									
Computation	18	.36	.50	.03	.55	.01	.35	.08	.23
Integers	27	.10	.32	- .17	.30	.20	.32	- .08	.27
Rational Numbers	13	- .13	.55	.13	.32	.11	.47	.09	.22
Percent	12	- .15	.46	- .12	.69	.12	.60	.04	.32
Ratio/Proportion	8	- .09	.20	.06	.40	.15	.37	- .08	.23
Average	10	- .41	.38	- .29	.42	.16	.44	- .18	.26
Denominate Numbers	12	- .47	.53	- .60	.49	- .57	.51	- .19	.31
Tables and Charts	8	.61	.34	- .76	.57	- .15	.30	- .24	.42
Algebra									
Algebraic Operations	18	.28	.43	.31	.60	.17	.46	.09	.37
Word Problems	38	- .24	.59	- .46	.67	- .14	.57	- .12	.34
Linear Functions	13	.21	.31	.09	.43	.02	.40	.03	.28
Quadratic Functions	9	.17	.20	.40	.29	.27	.34	.15	.28
Equations	12	.30	.36	.46	.58	.22	.22	.24	.23
Exponents	8	.07	.43	.26	.56	.00	.26	.11	.25
Miscellaneous	5	- .15	.31	- .68	.23	- .22	.23	- .20	.22
Geometry									
Points/Rays	10	.20	.30	.15	.58	- .34	.39	- .07	.25
Angles	9	- .07	.37	.48	.36	- .21	.29	.21	.22
Triangles	14	- .17	.51	.23	.52	.02	.44	.14	.25
Special Triangles	12	.08	.31	.51	.46	.26	.31	.25	.27
Circles	15	- .29	.47	.35	.53	.04	.43	.17	.29
Polygons	22	- .04	.38	.33	.55	.04	.42	.17	.28
3-Dimensional	8	- .59	.20	.23	.40	- .15	.36	.10	.36
Coordinate Geometry	8	- .29	.75	.39	.39	.04	.23	.23	.21
Miscellaneous									
Number System	10	.36	.39	.04	.62	.37	.24	.01	.32
Elementary Number System	13	.11	.20	- .29	.47	.06	.36	- .01	.19
Sets	11	.59	.39	- .11	.59	.18	.39	- .04	.42
Logic, Other	17	.17	.42	- .46	.75	- .07	.42	- .18	.39
F		5.28***		5.63***		2.74***		2.86***	

Note: Total number of SAT-M items = 360.

*p < .10. **p < .05. ***p < .01.

BEST COPY AVAILABLE

Table 11
Means and Standard Deviations for Item Categories on the SAT-M:
Points Tested (Spatial/Visual Factor)

Item Categories	Number of Items	Comparison							
		Male/Female		White/Asian		White/Black		White/Hispanic	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Spatial Factor									
Primary Spatial Component	9	-.17	.44	-.07	.28	-.22	.48	-.08	.30
No Figure But Helpful	18	-.19	.53	.06	.60	-.02	.47	.08	.26
Possible Spatial Factor	19	-.08	.48	.24	.59	-.17	.53	-.01	.33
Estimation Helpful	23	-.10	.48	.35	.45	.01	.46	.21	.22
No Spatial Factor	242	.04	.51	-.13	.65	.04	.46	-.03	.33
Ordinary Geometry	49	-.10	.49	.35	.48	.08	.34	.19	.29
F		1.72		7.79***		1.47		6.13***	
Stimulus Format									
Figure	65	-.11	.44	.29	.58	-.09	.46	.11	.31
Graph	11	-.47	.70	.01	.66	-.03	.20	.06	.34
Table	7	-.19	.39	-.35	.77	-.10	.37	-.07	.44
Number of Lines	7	-.05	.28	-.02	.28	-.11	.26	.01	.23
Venn Diagram	1	.84	-	-.50	-	-.71	-	-.70	-
Picture	1	.03	-	-.00	-	-.42	-	-.39	-
None	268	.03	.51	-.07	.63	-.07	.46	.00	.32
F		2.88***		3.23***		1.99*		2.22**	
Scale of Figure									
Not Drawn	7	-.20	.45	.39	.37	-.10	.30	.22	.27
Drawn to Scale	70	-.09	.44	.27	.55	-.08	.44	.09	.30
No Figure	283	.01	.53	-.08	.64	.05	.45	-.00	.33
F		1.72		10.31***		2.69*		3.84**	

Note: Total number of SAT-M items = 360.

*p < .10. **p < .05. ***p < .01.

Table 12
Means and Standard Deviations for Item Categories on the TSWE: Points Tested

Item Categories	Number of Items	Comparison							
		Male/Female		White/Asian		White/Black		White/Hispanic	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Item Type									
Usage	210	- .01	.39	- .09	.81	.01	.52	.01	.46
Sentence Correction	90	.08	.43	.18	.55	.08	.44	.08	.37
F		2.79*		8.14***		1.28		1.42	
Specification									
Subject/Verb Agreement	25	- .23	.29	- .23	1.01	- .29	.78	- .17	.35
Tense	18	.11	.52	- .18	.57	- .18	.30	.01	.41
Logic. Compar. & Agree.	11	- .12	.40	- .02	.39	- .05	.40	.28	.58
Pronouns	29	.06	.41	.09	.80	.00	.37	.20	.42
Idiom	37	- .01	.43	- .01	.67	.08	.45	.08	.41
Diction	11	- .07	.30	.22	.74	.12	.21	.11	.61
Usage Conventions	30	- .02	.34	.09	1.01	- .18	.65	- .26	.61
Shift	10	.24	.32	.57	.66	.12	.30	.28	.35
Sentence Boundaries	23	.10	.43	.19	.49	- .08	.45	.06	.29
Sentence Joining	25	.02	.48	- .02	.58	.11	.47	.02	.42
Dangling or Misplaced Modifier	12	- .05	.27	.23	.53	.14	.25	- .02	.28
Parallelism	17	.41	.31	.46	.52	.21	.34	.18	.29
Clarity/Economy	6	- .05	.45	.17	.39	.13	.47	.17	.25
No Error	46	- .02	.33	- .48	.57	.29	.38	- .02	.30
F		2.67***		3.28***		3.16***		2.78***	

Note: Total number of TSWE items = 300.

*p < .10. **p < .05. ***p < .01.

Table 13
Means and Standard Deviations for Item Content Categories on the SAT-V: Sentence Completions, Analogies, and Antonyms

Item Categories	Number of Items	Comparison: Male/Female Mean	S.D.	White/Asian Mean	S.D.	White/Black Mean	S.D.	White/Hispanic Mean	S.D.
Sentence Completions									
Aesthetics/Philosophy	23	.18	.40	- .19	.55	.03	.42	.08	.41
Practical Affairs	15	- .22	.74	- .11	.51	.22	.45	.11	.34
Science	19	- .36	.60	- .39	.40	- .25	.48	- .29	.43
Human Relationships	13	.55	.53	- .15	.53	.37	.42	.23	.31
Mixed/Overlapping	20	.13	.48	- .01	.66	.02	.45	.11	.25
F		6.38***		1.22		4.18***		4.96***	
Analogies									
Aesthetics/Philosophy	10	.26	.38	- .15	.50	.05	.55	.01	.62
Practical Affairs	22	- .12	.87	- .02	.56	- .14	.59	- .25	.57
Science	23	- .41	.82	- .41	.69	- .76	.50	- .67	.50
Human Relationships	24	.19	.46	.20	.43	.24	.28	.05	.38
Mixed/Overlapping	41	- .09	.64	- .01	.51	- .09	.61	- .09	.53
F		2.80**		3.82***		10.94***		6.79***	
Analogies -- Science									
Zoology/Biology	4	- .65	.56	.15	.54	- .56	.28	- .43	.47
Botany	2	.56	.55	- .52	.30	- .56	.07	- .78	.30
Physical Sci.	6	- .92	1.13	- .36	.68	- 1.04	.43	- .72	.45
Applied Sci.	9	- .23	.79	- .79	.45	.89	.50	- .73	.42
Math	1	.07	-	.07	-	.01	-	- .13	-
Mixed/Overlapping	6	- .43	.33	- .27	.67	- .60	.60	- .59	.58
No Science	92	.03	.63	.05	.51	.04	.53	- .06	.53
F		3.02***		4.19***		9.08***		4.61***	
Antonyms									
Aesthetics/Philosophy	10	.24	.23	.18	.39	.32	.30	.11	.43
Practical Affairs	22	- .44	.94	- .31	.65	- .01	.56	.04	.61
Science	16	- .37	1.01	- .24	.63	- .36	.82	- .21	.50
Human Relationships	37	.24	.75	.06	.54	.26	.44	.11	.50
Mixed/Overlapping	65	- .09	.61	- .05	.72	.20	.59	.00	.52
F		3.98***		1.78		4.18***		1.13	

*p < .10. **p < .05. ***p < .01.

Table 14
Means and Standard Deviations for Item Content Categories on the SAT-V: Reading Comprehension

Item Categories	Number of Items	Comparison							
		Male/Female		White/Asian		White/Black		White/Hispanic	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Content									
Humanities	24	.11	.36	- .14	.36	- .02	.37	.10	.31
Social Science	32	.12	.37	- .00	.34	.10	.44	.16	.29
Biology -- Traditional	5	- .52	.35	.35	.19	.11	.17	.37	.22
Biology -- General	8	- .41	.24	- .05	.32	.08	.33	.22	.27
Physical Science	22	- .07	.69	.05	.30	- .03	.26	.10	.28
Narrative	13	.63	.51	- .16	.31	- .04	.25	.13	.28
Argument: Humanities	22	.29	.37	- .15	.33	.14	.20	.21	.25
Argument: S.S.	10	.28	.61	- .09	.20	.04	.25	.07	.23
Argument: Science	11	- .03	.75	.12	.41	.06	.37	.18	.25
Argument: Mixed	3	.49	.16	- .15	.30	.52	.39	.44	.20
F		4.41***		1.97**		1.24		1.08	
Humanities Focus									
Literature Only	29	.25	.38	- .14	.35	.12	.34	.21	.22
Humanities	20	.16	.35	- .15	.33	.04	.33	.12	.37
None	101	.07	.61	.01	.33	.04	.34	.15	.28
F		1.24		3.18**		.55		.63	
Science Focus									
Science -- Technical	27	- .30	.60	.15	.30	.04	.28	.18	.29
Science -- General	19	.02	.66	- .02	.37	.01	.33	.15	.26
None	104	.24	.45	- .10	.33	.07	.36	.15	.29
F		12.38***		6.01***		.30		.14	
Structural Focus									
Descriptive	66	.05	.52	- .07	.31	.10	.41	.15	.30
Procedural	44	.16	.58	.04	.39	.00	.30	.14	.29
Argument/Persuasive	34	.27	.42	- .12	.31	.06	.24	.19	.25
Mixed	6	- .34	.81	.14	.44	- .08	.22	.19	.29
F		2.73**		2.09*		.97		.26	

*p < .10. **p < .05. ***p < .01.

Table 15
Means and Standard Deviations for Item Content Categories on the TSWE

Item Categories	Number of Items	Comparison							
		Male/Female		White/Asian		White/Black		White/Hispanic	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Arts	53	.01	.34	.00	.79	.06	.43	.02	.43
Social Science	56	-.06	.43	.25	.72	.06	.58	.12	.35
Science	52	-.09	.47	-.27	.71	-.11	.49	-.10	.43
Public Life	56	.01	.39	-.07	.79	.07	.48	-.01	.38
Student Relevant	31	.25	.37	.06	.75	.13	.42	.09	.65
Everyday Activ.	51	.09	.35	.01	.64	.04	.45	.10	.38
F		3.58***		2.77**		1.23		1.89*	

*p < .10. **p < .05. ***p < .01.

Table 16
Means and Standard Deviations for Item Content Categories on the SAT-M

Item Categories	Number of Items	Comparison: Male/Female Mean	S.D.	White/Asian Mean	S.D.	White/Black Mean	S.D.	White/Hispanic Mean	S.D.
Concrete/Abstract									
Real-Life (Word Problems)	90	-.36	.47	-.51	.66	-.20	.54	-.15	.36
Not Real-Life	270	.11	.47	.16	.53	.10	.39	.08	.30
F		66.04***		93.86***		30.64***		33.07***	
Relation to Curriculum									
Very Textbook-Like	65	.19	.50	.30	.66	.18	.44	.12	.30
Textbook-Like	270	-.05	.51	-.05	.61	-.00	.44	.00	.34
Not Textbook-Like	25	-.06	.42	-.36	.59	-.08	.49	-.05	.27
F		6.07***		12.59***		5.11***		4.33***	

*$p < .10$. **$p < .05$. ***$p < .01$.

Table 17
Means and Standard Deviations for Item Content Categories on the SAT-V: Technical/Non-Technical

Item Categories	Number of Items	Comparison							
		Male/Female		White/Asian		White/Black		White/Hispanic	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Sentence Completions									
Non-Technical	74	.08	.62	.16	.58	.06	.51	.02	.40
Stem Only	11	.05	.56	-.28	.45	-.00	.39	.08	.41
Key Only	0	-	-	-	-	-	-	-	-
Distractors Only	3	-.52	.61	-.03	.25	-.16	.38	.12	.30
Stem & Key Only	1	-.48	-	-.43	-	.02	-	.18	-
All/Mixed Tech	1	-.84	-	.25	-	.35	-	.29	-
F		1.36		.34		.27		.21	
Reading Comprehension									
Technical (P & Q)	19	-.31	.60	.14	.29	.03	.26	.26	.23
Passage Tech	25	-.03	.50	-.06	.36	.02	.29	.06	.34
Question Tech	2	-.25	1.36	.51	.16	-.25	.17	.37	.07
General (P & Q)	104	.24	.47	-.08	.33	.07	.37	.16	.27
F		7.09***		4.22***		.72		2.11*	
Analogies									
Non-Technical	92	.04	.60	-.04	.55	-.10	.60	-.17	.56
Stem Only	7	-.51	.82	-.48	.60	-.77	.60	-.59	.53
Key Only	6	-.20	.35	.26	.47	.17	.49	.02	.50
Distractors Only	8	-.12	.95	-.30	.60	-.14	.53	-.18	.55
Stem & Key Only	2	-1.24	1.20	.23	.14	.18	.44	.20	.34
All/Mixed Tech	5	-.93	.71	.13	.76	-.65	.52	-.56	.60
F		4.07***		1.63		2.71**		1.47	
Antonyms									
Non-Technical	139	-.05	.79	-.06	.66	.15	.60	.02	.52
Stem Only	4	-.02	.42	.09	.64	-.04	.83	.01	.99
Key Only	0	-	-	-	-	-	-	-	-
Distractors Only	6	-.57	.53	-.34	.58	-.07	.39	.04	.35
Stem & Key Only	0	-	-	-	-	-	-	-	-
All/Mixed Tech	1	-.52	-	-.07	-	-.10	-	.10	-
F		.99		.42		.41		.01	

*p < .10. **p < .05. ***p < .01.

Table 18
Means and Standard Deviations for Item Content Categories on the Analogies Section of the SAT-V: Concrete/Abstract

Item Categories	Number of Items	Comparison							
		Male/Female		White/Asian		White/Black		White/Hispanic	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Analogies									
Concrete	40	-.22	.84	-.47	.60	-.70	.60	-.61	.60
Mixed	48	-.11	.67	.13	.48	.09	.44	-.06	.43
Abstract	32	-.16	.49	.18	.39	.18	.37	.12	.40
F		2.64*		19.31***		39.12***		22.59***	

*p < .10. **p < .05. ***p < .01.

Table 19
Means and Standard Deviations for Item Content Categories on the SAT-V: Latinate Language

Item Categories	Number of Items	Comparison							
		Male/Female		White/Asian		White/Black		White/Hispanic	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Analogies									
No Latinate	20	-.02	.90	-.54	.68	-.50	.62	-.70	.55
All Latinate	19	.01	.43	.30	.45	.13	.56	.14	.53
Mixed	81	-.10	.71	-.02	.50	-.13	.60	-.15	
F		.24		12.54***		5.41***		13.81***	
Antonyms									
No Latinate	8	.13	.83	-.24	.68	.04	.53	-.07	.46
All Latinate	89	-.10	.76	.03	.61	.13	.56	.09	.54
Mixed	53	-.05	.78	-.20	.70	.15	.66	-.08	.52
F		.34		2.27		.10		1.90	

*p < .10. **p < .05. ***p < .01.

Table 20
Means and Standard Deviations for Item Content Categories on the Analogies and Antonyms Sections of the SAT-V: Homographs

Item Categories	Number of Items	Comparison							
		Male/Female		White/Asian		White/Black		White/Hispanic	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Analogies -- Homographs									
In Stem									
One	39	-.06	.66	-.14	.59	-.18	.56	-.13	.57
Two	8	.38	.54	-.57	.31	-.62	.56	-.82	.43
None	73	-.13	.73	.04	.56	-.08	.63	-.16	.55
F		1.90		4.87***		2.85*		5.43***	
In Key									
One	37	-.03	.78	-.17	.60	-.19	.70	-.30	.58
Two	5	.20	.21	-.34	.27	-.27	.38	-.23	.56
None	78	-.11	.69	.01	.57	-.12	.59	-.14	.56
F		.52		1.89		.26		.92	
In Distractor									
One or More	61	-.04	.76	-.20	.59	-.31	.66	-.35	.61
None	59	-.11	.66	.09	.53	.02	.52	-.03	.48
F		.23		7.52***		8.97***		9.65***	
Antonyms -- Homographs									
Stem Only	11	.04	.33	-.26	.60	.11	.27	.07	.57
Key Only	11	-.01	.64	.08	.57	.32	.44	-.00	.26
Distractor(s) Only	32	-.15	.92	-.14	.82	.08	.72	-.03	.59
Stem or Key	3	.48	.33	-.32	.61	-.29	.13	-.44	.31
Stem/Key/Distractor(s)	28	-.26	.85	.04	.43	.16	.54	.13	.58
None	65	.00	.72	-.06	.65	.14	.62	.02	.50
F		.88		.61		.57		.77	

*$p < .10$. **$p < .05$. ***$p < .01$.

BEST COPY AVAILABLE

Table 21
Means and Standard Deviations for Item Content Categories on the SAT-V: Parts of Speech

Item Categories	Number of Items	Comparison							
		Male/Female		White/Asian		White/Black		White/Hispanic	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Sentence Completions (In Options)									
One Noun	18	.09	.52	-.14	.69	.25	.41	.20	.25
One Adjective	20	-.04	.85	-.18	.62	-.17	.52	-.08	.52
One Verb	7	-.14	.47	-.60	.18	-.18	.35	-.12	.39
Two Nouns	5	.08	.45	.03	.49	.52	.42	.21	.25
Two Adjectives	14	-.07	.48	-.16	.42	-.14	.41	-.02	.26
Two Verbs	4	-.07	.22	-.48	.44	.27	.20	-.13	.20
Mixed	22	.22	.62	-.06	.46	.14	.45	.08	.42
F		.55		1.21		3.25***		1.42	
Analogies									
Noun: Noun	58	-.15	.83	-.19	.60	-.39	.66	-.33	
Noun: Adjective/Adverb	26	-.08	.59	.11	.43	.04	.54	-.00	.43
Noun: Verb	24	-.01	.50	.01	.68	.04	.45	-.13	.60
Verb: Verb	3	-.47	.42	-.37	.41	-.20	.03	-.70	.17
Verb: Adjective	7	.36	.38	.15	.22	.41	.27	.16	.39
Adjective: Adjective	2	.61	.67	.51	.11	.34	.44	.13	.03
F		1.26		1.82		4.74***		2.62**	
Antonyms									
Verb	49	-.19	.80	-.22	.68	.19	.65	-.11	.51
Noun	41	.02	.70	.12	.52	.12	.66	.08	.55
Adjective/Adverb	60	-.04	.79	-.07	.69	.08	.50	.09	.52
F		.88		3.02**		.44		2.24	

*p < .10 **p < .05 ***p < .01

Table 22
Means and Standard Deviations for the Emotive Item Content Category on the SAT-V and TSWE

Item Categories	Number of Items	Comparison							
		Male/Female		White/Asian		White/Black		White/Hispanic	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Emotive									
Analogies									
Strong	8	-.12	.43	-.09	.80	-.42	.57	-.60	.62
Neutral	112	-.07	.73	-.06	.56	-.13	.62	-.17	.56
F		.04		.03		1.59		4.30**	
Antonyms									
Strong	4	-.46	.50	-.23	.68	.25	1.35	-.62	.66
Neutral	146	-.06	.78	-.06	.66	.13	.56	.04	.52
F		1.05		.26		.15		6.18**	
Sentence Completions									
Strong	4	-.52	.51	-.25	.46	-.16	.44	.02	.34
Neutral	86	.07	.62	-.17	.56	.06	.49	.04	.40
F		3.50*		.09		.75		.01	
Reading Comprehension									
Strong	8	.26	.61	-.23	.39	.08	.53	.14	.33
Neutral	142	.11	.54	-.03	.34	.05	.33	.16	.28
F		.53		2.37		.05		.02	
TSWE									
Strong	8	-.08	.17	-.03	.56	.07	.35	.08	.16
Neutral	292	.02	.41	-.01	.76	.03	.50	.03	.44
F		.46		.00		.05		.10	

*p < .10. **p < .05. ***p < .01.

Table 23
Means and Standard Deviations for Item Content Categories on the SAT-V and TSWE: Controversial and Socially Relevant

Item Categories	Number of Items	Comparison							
		Male/Female		White/Asian		White/Black		White/Hispanic	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Controversial									
Sentence Completions									
Yes	14	-.32	.45	-.10	.45	.08	.47	-.01	.35
No	76	.11	.63	-.19	.57	.05	.49	.05	.41
F		5.92**		.27		.07		.26	
Analogies									
Yes	17	-.37	.42	.14	.47	.11	.34	.04	.37
No	103	-.03	.74	-.09	.59	-.19	.65	-.23	.59
F		3.44*		2.33		3.55*		3.33*	
TSWE									
Yes	70	-.10	.40	.03	.76	-.02	.48	.03	.37
No	230	.05	.40	-.02	.75	.05	.50	.03	.46
F		7.43***		.29		1.08		.00	
Socially Relevant									
Sentence Completions									
Yes	16	-.19	.40	-.14	.44	.19	.46	.04	.34
No	74	.09	.65	-.18	.58	.02	.49	.04	.41
F		2.81*		.05		1.55		.01	
Reading Comprehension									
Yes	8	.26	.61	-.23	.39	.08	.53	.14	.33
No	142	.11	.54	-.03	.34	.05	.33	.16	.28
F		.53		2.37		.05		.02	

*p < .10. **p < .05. ***p < .01.

Table 24
Means and Standard Deviations for the Minority Stimulus Item Content Categories on the Sentence Completions and Reading Comprehension Sections of the SAT-V

Item Categories	Number of Items	Comparison: Male/Female Mean	Male/Female S.D.	White/Asian Mean	White/Asian S.D.	White/Black Mean	White/Black S.D.	White/Hispanic Mean	White/Hispanic S.D.
Sentence Completions--References									
Black American	3	.31	.25	.17	.59	.16	.12	.04	.26
Asian American	1	-.31	-	.51	-	.18	-	-.23	-
Third World Black	1	.62	-	-.46	-	.17	-	-.69	-
Non-Minority	5	.00	.46	-.06	.49	.20	.41	.17	.14
Unidentifiable	56	.21	.50	-.17	.56	.10	.46	.08	.38
No People	24	-.39	.74	-.27	.53	-.11	.57	-.05	.45
F		4.03***		.74		.75		1.26	
Reading Comprehension--References									
Black American	10	.24	.39	-.18	.24	.39	.30	.22	.16
Hispanic American	5	.22	.31	.37	.29	-.27	.30	-.11	.30
Native American	15	.14	.41	-.10	.30	.09	.44	.23	.29
Non-Minority	5	.08	.22	.03	.13	.16	.28	.30	.28
Unidentifiable	100	.19	.54	-.05	.36	.05	.31	.16	.29
No People	15	-.51	.55	-.03	.34	-.10	.33	.06	.21
F		5.03***		1.86*		3.81***		1.75	
Sentence Completions--Named									
Black American	2	.49	.04	.47	.50	.23	.09	.19	.20
Asian American	1	-.31	-	.51	-	.18	-	.23	-.46
Non-Minority	1	.25	-	.56	-	.75	-	.40	-.74
Unidentifiable	15	.47	.33	-.19	.59	.13	.46	.07	.41
No One Named	71	-.06	.64	-.21	.53	.02	.49	.02	.40
F		2.73**		1.57		.78		.42	
Reading Comprehension--Named									
Black American	6	.20	.44	-.17	.31	.49	.33	.29	.17
Native American	2	.61	.01	-.21	.05	.35	.16	.29	.12
Non-Minority	8	.07	.23	.11	.21	.17	.23	.33	.25
Unidentifiable	57	.23	.45	.01	.32	.05	.29	.20	.26
No One Named	77	.02	.62	-.09	.37	.01	.36	.09	.30
F		1.72		1.36		3.58***		2.53**	

Note: All Sentence Completions and Reading Comprehension items were coded using the following categories: Black American, Hispanic American, Native American, Asian American, Third World Black, Third World Hispanic, Third World Asian, Non-Minority, or No People; only cells with items were included above.

*p < .10. **p < .05. ***p < .01.

Table 25
Means and Standard Deviations for the Gender Reference Item Content Categories on the Sentence Completions and Reading Comprehension Sections of the SAT-V

Item Categories	Number of Items	Comparison							
		Male/Female		White/Asian		White/Black		White/Hispanic	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Sentence Completions--Reference									
Female Only	17	.27	.51	.04	.53	.31	.40	.17	.27
Male Only	14	.39	.39	.00	.55	.11	.51	.15	.30
Mixed	3	.46	.26	.43	.39	.34	.06	.11	.22
Unidentifiable	32	.06	.49	- .35	.51	- .02	.40	- .03	.43
No People	24	- .39	.74	- .27	.53	- .11	.57	- .05	.45
F		5.82***		3.04**		2.47**		1.31	
Reading Comprehension-Reference									
Female Only	0	-	-	-	-	-	-	-	-
Male Only	40	.28	.40	.03	.27	.21	.33	.25	.26
Mixed	27	.41	.42	- .14	.33	.00	.32	.13	.24
Unidentifiable	68	.04	.54	- .02	.38	.02	.33	.14	.31
No People	15	- .51	.55	- .03	.34	- .10	.33	.06	.21
F		13.37***		.88		4.19***		2.20*	
Sentence Completions-Named									
Female Only	4	.48	.15	- .19	.28	.21	.22	.22	.35
Male Only	10	.44	.37	- .06	.64	.12	.57	.10	.32
Mixed	2	.61	.17	.54	.43	.30	.02	.13	.26
Unidentifiable	3	.12	.36	- .14	.74	.25	.19	- .16	.55
No One Named	71	- .06	.64	- .21	.53	.02	.50	.02	.40
F		2.55**		.98		.48		.50	
Reading Comprehension-Named									
Female Only	1	.60	-	- .26	-	.50	-	.41	-
Male Only	58	.22	.46	.02	.29	.11	.31	.23	.26
Mixed	14	.22	.32	- .06	.38	.05	.31	.20	.26
Unidentifiable	0	-	-	-	-	-	-	-	-
No One Named	77	.02	.62	- .09	.37	.01	.36	.09	.30
F		1.90		1.15		1.53		2.85**	

*p < .10. **p < .05. ***p < .01.

Table 26
Means and Standard Deviations for the Gender Reference and Minority Stimulus Item Content Categories on the TSWE

Item Categories	Number of Items	Comparison							
		Male/Female		White/Asian		White/Black		White/Hispanic	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Gender Reference									
Female Only	34	.13	.34	-.16	.62	.07	.42	.13	.38
Male Only	63	.07	.42	.01	.79	.08	.49	.08	.46
Mixed	13	.11	.54	-.16	.92	.02	.60	-.18	.45
Unidentifiable	131	.04	.38	.07	.77	.05	.51	.05	.41
No People	59	-.17	.40	-.09	.67	-.08	.48	-.09.	.47
F		4.38***		1.03		.99		2.58**	
Minority Stimulus									
Black American	18	.10	.45	.16	.70	.20	.51	.15	.45
Hispanic American	4	-.17	.22	-.57	.32	-.27	.45	-.25	.24
Native American	5	.02	.40	.03	.50	-.08	.62	.06	.41
Asian American	-	-	-	-	-	-	-	-	-
Third World Black	4	-.32	.11	.10	.53	-.17	.50	.02	.09
Third World Hispanic	2	-.16	.06	-.18	.52	.29	.02	.32	.01
Third World Asian	2	-.16	.08	.11	.76	.25	.01	-.11	.03
Non-Minority	11	-.02	.38	.07	.60	.19	.39	-.09	.31
Unidentifiable	196	.08	.40	.01	.79	.05	.50	.07	.43
No People	58	-.18	.39	-.09	.67	-.09	.48	-.09	.48
F		3.24***		.54		1.28		1.33	

*p < .10. **p < .05. ***p < .01.

Table 27
Means and Standard Deviations for the People Reference Item Content Categories on the Analogies and Antonyms Sections of the SAT-V

Item Categories	Number of Items	Comparison							
		Male/Female		White/Asian		White/Black		White/Hispanic	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Analogies									
People Reference	23	-.18	.58	.27	.41	.14	.46	.12	.48
No People	97	-.05	.74	-.13	.59	-.22	.64	-.27	.57
F		.66		9.37***		6.47***		8.75***	
Antonyms									
People Reference	5	-.47	1.19	.16	.40	.22	.71	-.11	.81
No People	145	-.06	.75	-.07	.66	.13	.59	.02	.52
F		1.39		.60		.11		.32	

Note: None of the People References in the Analogies and Antonyms sections of the SAT-V was identifiable by gender or minority group.

*p < .10. **p < .05. ***p < .01.

Table 28
Means and Standard Deviations for the Gender Reference and Minority Stimulus Item Content Categories on the SAT-M

Item Categories	Number of Items	Comparison							
		Male/Female		White/Asian		White/Black		White/Hispanic	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Gender Reference									
Female	6	-.39	.45	-.43	.60	-.32	.51	-.18	.29
Male	6	-.11	.25	-.55	.63	-.07	.62	-.15	.19
Mixed	33	-.28	.55	-.70	.72	-.19	.61	-.24	.42
No People	315	.03	.50	.09	.57	.06	.42	.05	.30
F		4.90***		20.47***		4.27**		10.08***	
Minority Stimulus									
Black	0	-	-	-	-	-	-	-	-
Hispanic	1	.73	-	-1.53	-	-.80	-	-.47	-
Other Minority	0	-	-	-	-	-	-	-	-
Unidentifiable People	44	-.29	.49	-.63	.69	-.18	.60	-.22	.38
No People	315	-.03	.50	.09	.57	.06	.42	.05	.30
F		8.99***		31.42***		6.92***		15.21***	

*p < .10. **p < .05. ***p < .01.

Table 29
Summary of Analysis of Variance Results for the Item Format Categories Dealing with Length on the SAT-V and TSWE

	Male/Female	White/Asian	White/Black	White/Hispanic
Sentence Completions				
Length of Stem	-	.010[a]	-	-
Length of Options	-	-	-	-
Reading Comprehension				
Length of Stimulus - 5 Categories	-	-	-	.070[c]
Length of Stimulus - 2 Categories	-	-	-	.060[a]
Length of Stem	-	-	-	-
Length of Options	.061[c]	-	-	-
Opt.: Short Phrase/Long Phrase/Sent.	-	-	-	-
TSWE - Usage				
Length of Stem	-	-	-	
Length of Options	.067[b]	-	-	-
TSWE - Sentence Correction				
Length of Stem	-	-	-	-
Length of Options	-	-	-	.098[c]

Note: A dash indicates that the F value was not significant at an alpha level of .10.

[a] Increased length favored focal group.
[b] Increased length favored reference group.
[c] No consistent relationship between length and reference/focal group performance difference.

Table 30
Summary of Analysis of Variance Results for the Item Format Categories Dealing with Length on the SAT-M

	Male/Female	White/Asian	White/Black	White/Hispanic
Regular Mathematics				
Length of Stimulus	.10[a]	-	-	-
Length of Stem	.00[b]	.00[b]	.00[b]	.00[b]
Quantitative Comparison				
Length of Stimulus	.00[b]	.00[b]	-	.06[c]
Length of Stem	.00[b]	.01[b]	-	-

Note: A dash indicates that the F value was not significant at an alpha level of .10.

[a] Increased length favored focal group.
[b] Increased length favored reference group.
[c] No consistent relationship between length and reference/focal group performance difference.

Table 31
Means and Standard Deviations for the Option Format Item Category on the SAT

Item Categories	Number of Items	Comparison: Male/Female Mean	Male/Female S.D.	White/Asian Mean	White/Asian S.D.	White/Black Mean	White/Black S.D.	White/Hispanic Mean	White/Hispanic S.D.
Analogics									
Options Listed Vertically	28	.07	.54	.11	.38	.10	.41	.16	.45
Options Listed Horizontally	92	-.12	.75	-.11	.62	-.23	.65	-.30	.56
F		1.50		2.96**		6.02**		15.91***	
Sentence Completions									
Options Listed Vertically	33	.15	.48	-.08	.40	.10	.43	.05	.27
Options Listed Horizontally	57	-.02	.69	-.22	.62	.02	.51	.03	.45
F		1.52		1.35		.58		.06	
Reading Comprehension									
Options Listed Vertically	135	.09	.55	-.04	.35	.06	.35	.16	.29
Options Listed Horizontally	15	.32	.52	-.07	.29	.05	.32	.14	.28
F		2.39		.10		.00		.09	
Mathematics									
Options Listed Vertically	122	.01	.53	-.06	.67	.03	.51	.01	.34
Options Listed Horizontally	118	-.03	.51	.02	.70	-.01	.43	.04	.35
NA	120	-.01	.50	.03	.53	.05	.42	.01	.29
F		.16		.66		.56		.34	
Math--Alternate Coding									
Options Listed Vertically	122	.01	.53	-.06	.67	.03	.51	.01	.34
Options Listed Horizontally (1 line)	49	-.17	.54	-.01	.74	-.03	.44	.05	.40
Options Listed Horizontally (wrap)	69	.07	.46	.03	.66	.01	.42	.03	.31
NA	120	-.01	.50	.03	.53	.05	.42	.01	.29
F		2.19*		.48		.43		.27	

Note: Typically the options for Antonym items are listed horizontally; only one Antonym item had options listed vertically.

*$p < .10$. **$p < .05$. ***$p < .01$.

Table 32
Means and Standard Deviations for the Item Format Categories Dealing with First Appearance and First Reappearance of an Item Type on the SAT-V and TSWE

Item Categories	Number of Items	Comparison							
		Male/Female		White/Asian		White/Black		White/Hispanic	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Verbal									
First Appearance	24	-.13	.92	- .30	.51	-.38	.57	-.26	.53
Reappearance-1st Sect.	4	.23	.37	- .46	.36	-.29	.33	-.27	.31
First Appearance-2nd Sect.	24	.03	.55	- .28	.68	-.24	.59	-.29	.58
Reappearance-2nd Sect.	2	.10	.31	- .27	.18	-.06	.05	-.13	.20
Other items	456	.01	.67	- .05	.54	.07	.52	.05	.46
F		.37		2.74**		6.25***		5.55***	
Verbal--Alternate Coding									
First Appr./Reappr. Combined	54	-.02	.73	- .30	.58	-.30	.56	-.27	.54
Other Items	456	.22	.67	- .05	.54	.07	.52	.05	.46
F		.10		10.64***		23.83***		22.09***	
TSWE									
First Appearance	12	.07	.45	.03	1.00	-.02	.87	-.13	.34
Reappearance	6	.17	.38	- .00	.71	-.12	.47	.25	.24
Other Items	282	.01	.41	- .01	.74	.04	.48	.03	.44
F		.52		.00		.36		1.56	
TSWE--Alternate Coding									
First Appr./Reappr. Combined	18	.10	.43	- .02	.91	-.05	.76	-.06	.36
Other items	282	.01	.41	- .01	.74	.04	.48	.03	.44
F		.80		.001		.54		.11	

*p < .10. **p < .05. ***p < .01.

BEST COPY AVAILABLE

Table 33
Means and Standard Deviations for the Item Format Categories Dealing with First Appearance and First Reappearance of an Item Type on the SAT-M

Item Categories	Number of Items	Comparison							
		Male/Female		White/Asian		White/Black		White/Hispanic	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Mathematics									
First Appearance	7	.31	.34	- .12	.45	-.11	.24	.03	.24
Reappearance-1st Sect.	1	.31	-	.27	-	.16	-	.26	-
First Appearance-2nd Sect.	11	.15	.43	- .01	.49	-.02	.41	.10	.21
Reappearance-2nd Sect.	5	.01	.53	- .16	.98	-.24	.92	-.04	.47
Other items	336	- .02	.51	- .00	.64	.03	.45	.02	.33
F		3.36*		.18		.67		.34	
Math--Alternate Coding									
First Appr./Reappr. Combined	24	.17	.43	- .06	.61	-.09	.53	.06	.30
Other Items	336	- .02	.51	- .00	.64	.03	.45	.02	.33
F		3.36*		.21		1.62		.33	

*$p < .10$. **$p < .05$. ***$p < .01$.

Table 34
Means and Standard Deviations for Item Format Categories on the Analogies Section of the SAT-V: Clues to the Answer

Item Categories	Number of Items	Comparison							
		Male/Female		White/Asian		White/Black		White/Hispanic	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Idea Association									
Stem and Key									
Vertical Rel.	60	- .12	.77	- .15	.59	- .30	.66	- .35	.58
No Vertical Rel.	60	- .03	.64	.03	.55	- .00	.55	- .04	.52
F		.47		2.95*		7.06***		9.56***	
Stem and Distract.									
Vertical Rel.	74	- .17	.77	- .12	.63	- .27	.68	- .31	.63
No Vertical Rel.	46	.09	.56	.04	.48	.05	.46	- .01	.41
F		3.89***		2.13		7.89***		7.76***	
Independent/Overlap									
Independent	80	- .10	.71	.01	.54	- .05	.56	- .09	.50
Overlapping	40	- .02	.71	- .20	.63	- .36	.69	- .40	.64
F		.34		3.81**		6.82***		8.26***	

*$p < .10$. **$p < .05$. ***$p < .01$.

Table 35
Means and Standard Deviations for Item Format Categories on the Sentence Completions and Reading Comprehension Sections of the SAT-V: Clues to the Answer

Item Categories	Number of Items	Comparison							
		Male/Female		White/Asian		White/Black		White/Hispanic	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Sentence Completions									
Means of Finding Answer									
After reading Q.	66	.09	.58	- .16	.57	.00	.51	.02	.42
Look at Opt.	24	- .10	.71	- .20	.50	.18	.38	.08	.32
F		1.63		.08		2.33		.43	
Reading Comprehension									
Means of Finding Answer									
After Reading Q.	54	.08	.60	- .09	.37	.08	.40	.12	.31
Look at Opt.	96	.14	.52	- .02	.33	.04	.31	.18	.26
F		.44		1.31		.37		1.42	
Line reference									
Specific Line	10	.17	.32	- .30	.33	-.02	.31	.19	.28
2-4 Lines	14	.31	.36	- .15	.24	.05	.46	.12	.23
Specific Paragraph	5	- .07	.42	- .18	.34	.09	.21	.20	.17
No Line Reference	121	.10	.58	- .00	.34	.06	.33	.16	.29
F		.86		3.21**		.18		.18	
Specific/Non-Specific									
Stem/Opt. Specific	82	.04	.58	- .06	.38	- .06	.39	.17	.29
Opt. Specific	30	.16	.48	.01	.31	.13	.29	.20	.27
Stem Specific	25	.41	.38	.10	.30	- .02	.25	.08	.25
Stem/Opt. Not Specific	13	- .02	.56	- .09	.25	- .00	.26	.12	.30
F		3.36**		1.22		1.02		1.04	

*p < .10. **p < .05. ***p < .01.

Means and Standard Deviations for Item Format Categories of the SAT-M

Item Categories	Number of Items	Comparison							
		Male/Female		White/Asian		White/Black		White/Hispanic	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Reading Difficulty									
Difficult	26	- .06	.53	- .39	.71	- .04	.57	- .13	.42
Medium	245	- .08	.52	- .06	.64	- .01	.46	.01	.33
Easy	89	.19	.43	.25	.49	.14	.36	.08	.26
F		9.84***		13.76***		3.75**		4.12**	
Reading Level									
Difficult	74	- .07	.55	- .38	.66	- .07	.51	- .10	.37
Medium/Easy	286	.01	.50	.09	.59	.05	.43	.05	.31
F		1.42		34.80***		4.03**		12.81***	
"Cannot be determined" Option									
Attribute present	13	- .23	.41	- .21	.74	- .00	.55	- .03	.36
Attribute absent	227	- .00	.52	- .01	.68	.01	.46	.03	.34
N/A	120	- .01	.50	.03	.53	.05	.42	.01	.29
F		1.24		.80		.39		.33	
Max/Min									
Attribute present	14	- .16	.41	- .43	.69	- .27	.37	- .34	.28
Attribute absent	226	- .00	.52	.01	.68	.03	.47	.05	.34
N/A	120	- .01	.50	.02	.53	.05	.42	.01	.29
F		.58		3.29**		3.21**		9.72***	
Must/Could									
"Must" in Stem	14	- .00	.28	- .17	.80	- .09	.51	- .08	.49
"Could" in Stem	15	.08	.48	- .04	.63	.19	.49	.11	.33
N/A	331	- .01	.52	.00	.63	.02	.45	.02	.32
F		.23		.53		1.42		1.30	
Role of Options									
Independent of Option	188	- .03	.52	- .04	.72	- .03	.47	.01	.35
Dependent on Option	52	.04	.49	.06	.53	.17	.43	.10	.31
N/A	120	- .01	.50	.03	.53	.05	.42	.01	.29
F		.35		.76		4.75***		1.72	
Order of Options									
Least to Greatest	147	- .07	.52	- .13	.70	- .06	.48	- .03	.34
Greatest to Least	35	- .03	.50	.07	.62	.03	.48	.08	.29
N/A	178	.04	.50	.08	.57	.09	.42	.05	.32
F		1.77		4.64***		4.79***		2.68*	
Underlining in Stem									
Attribute Present	10	- .30	.38	- .23	.33	.04	.34	.17	.31
Attribute Absent	350	- .00	.51	.00	.64	.02	.46	.02	.33
F		3.25*		1.29		.01		2.05	

Note: Total number of SAT-M items = 360.

*p < .10. **p < .05. ***p < .01.

BEST COPY AVAILABLE

Table 37
Means and Standard Deviations for Item Format Categories on the TSWE

Item Categories	Number of Items	Comparison							
		Male/Female		White/Asian		White/Black		White/Hispanic	
		Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
Sentence Structure									
Simple	71	-.01	.39	-.05	.77	-.00	.47	-.01	.40
Compound	25	.06	.42	.02	.68	-.12	.42	.08	.37
Complex/1 Dependent Clause	174	-.01	.41	.03	.76	.04	.53	.03	.47
Complex/2+ Dependent Clauses	28	.16	.37	-.15	.67	.18	.40	.07	.38
Compound/Complex	2	.53	.44	.13	.79	.45	.10	-.04	.15
F		1.99*		.44		1.67		.27	
Key: Error/No Error									
Error	241	.04	.42	.07	.76	-.02	.51	.04	.46
No Error	59	-.07	-.36	-.33	.63	.25	.38	-.02	.31
F		3.55*		14.40***		14.96***		.94	
Underline Position									
Including First Word	11	.07	.46	.18	.59	.13	.41	.08	.38
Not First or Last Word	40	.03	.42	.15	.52	-.01	.46	.05	.34
Including Last Word	29	.07	.43	.17	.61	.11	.42	.03	.32
Entire Sentence	10	.27	.42	.27	.44	.23	.22	.16	.24
N/A	210	-.01	.39	-.09	.81	.01	.52	.02	.48
F		1.39		1.94*		.85		.32	

Note: Total number of TSWE items = 300.

*p < .10. **p < .05. ***p < .01.

Table 38
Summary of Analysis of Variance Results for Selected
Item Format Categories on the SAT

	Male/Female	White/Asian	White/Black	White/Hispanic
Key				
Mathematics				
Quantitative Comparison	-	-	-	-
Regular Mathematics	-	-	-	-
Verbal				
Reading Comprehension	-	-	.010[a]	.015[a]
Analogies	-	-	-	-
Sentence Completions	-	-	-	-
Antonyms	.079[b]	-	-	-
Roman Numeral Format				
Regular Mathematics	-	-	-	-
Reading Comprehension	-	-	-	-
Negative Item				
Mathematics	.098[c]	-	-	-
Reading Comprehension	-	-	-	-
Open-Closed Stem				
Reading Comprehension	-	-	-	-

Note: A dash indicates that the F value was not significant at an alpha level of .10.

[a] Items keyed with option "B" tended to favor Whites; Options "A," "C," and "E" tended to favor focal groups.
[b] Items keyed with options "B" and "D" tended to favor males; options "A" and "E" tended to favor females.
[c] Items with stems stated in the negative (number of items = 16) tended to favor males.